Numerical impacts on tracer transport: A proposed intercomparison test of Atmospheric General Circulation Models

Aman Gupta[1*], Edwin P. Gerber[1] and Peter H. Lauritzen[2]

[1]*Center for Atmosphere-Ocean Sciences, Courant Institute of Mathematical Sciences, New York University, NY, USA*

[2]*National Center for Atmospheric Research, Boulder, Colorado*

[*]Correspondence to: agupta@cims.nyu.edu

The transport of trace gases by the atmospheric circulation, radiatively active species and their chemical precursors, plays an important role in the climate system and its response to external forcing. Transport presents a challenge for Atmospheric General Circulation Models (AGCMs), where errors in both the resolved circulation and the numerical representation of trace gases can bias transport. In this study, two benchmark tests are established to assess transport by the dynamical core of an AGCM. To separate transport from chemistry, the tests focus on the age-of-air, an estimate of the mean transport time by the circulation. The tests assess the coupled stratosphere-troposphere system, focusing on the overturning circulation of the stratosphere, or Brewer-Dobson Circulation, where transport time scales on the order of months to years provide a stiff test of model numerics.

To establish a benchmark, four dynamical cores employing different numerical schemes (finite-volume, pseudo-spectral, and spectral-element) and discretizations (cubed sphere vs. latitude-longitude) are compared across a range of resolutions. The subtle momentum balance of the tropical stratosphere is very sensitive to model numerics, and the first intercomparison reveals stark differences in climatological circulation, particularly at high vertical resolution; some cores develop westerly jets and others, easterly jets. This leads to substantial spread in transport, biasing the age-of-air by up to 25% relative to its climatological mean, making it difficult to assess the impact of the numerical representation of trace gases on transport. This uncertainty is remove by constraining the tropical winds in the second intercomparison test, akin to specifying the Quasi-Biennial Oscillation in an AGCM. The dynamical cores exhibit qualitative agreement on the structure of atmospheric transport in the second test, with evidence of convergence as the horizontal and vertical resolution is increased. Significant quantitative differences remain, however, particularly between models employing spectral vs. finite-volume numerics, even in state-of-the-art cores.

Key Words: Tracer Transport, Stratospheric Dynamics, Dynamical Cores, Age of air, Brewer-Dobson Circulation

Received . . .

1. Introduction

It has been a quarter of a century since Held and Suarez (1994) proposed a test to compare the dynamical cores of atmospheric general circulation models, the numerical solvers that integrate the primitive equations on the sphere. The Held and Suarez 1994 test, hereafter referred to as HS94, is a simple recipe for parameterizing all non-conservative processes for a dry (moisture free) atmosphere, including diabatic processes, such as radiative transfer and convection, and frictional processes, specifically the atmospheric boundary layer. This allowed them to see if two dynamical cores with very different numerics — a pseudo-spectral vs. a finite-difference scheme — would produce the same climatological circulation.

The HS94 benchmark was designed to compliment existing tests for dynamical cores, which examine the evolution of numerics on shorter time scales in more controlled settings, e.g., the reproduction of a single baroclinic life cycle experiment (Williamson et al. 1992). In the ensuing decades, however, the HS94 recipe for producing a fairly realistic climate with such a simple parameterization of diabatic processes has lead to substantial amount of research on the dynamics and circulation of the atmosphere (e.g., Son and Lee 2005; Lorenz and DeWeaver 2007; Butler et al. 2010).

In this paper, we propose an update of the HS94 benchmark, designed to take into account two advances in climate research over the past 25 years. First, there has been a growing awareness that the tropospheric circulation is tightly coupled with that of the stratosphere, such that climate and weather prediction is improved by accurately representing the stratosphere (e.g., Gerber et al. 2012). We update the HS94 recipe to allow one to explore the impact of numerics and resolution on the coupled stratosphere-troposphere circulation. Following the work of Polvani and Kushner (2002), the test allows to compare the circulation and variability of the stratosphere, Sudden Stratospheric Warming (SSW) events in particular.

Second, atmospheric chemistry has become an important element in climate prediction, as evidenced by the AerChemMIP (Collins et al. 2017) within the Coupled Model Intercomparsion Project, Phase 6 (CMIP6; Eyring et al. 2016). Our test seeks to ensure that dynamical cores consistently represent the transport of trace gases by the resolved atmospheric flow. Transport – the advection, mixing, and diffusion of constituents by the atmospheric circulation – plays a fundamental role in setting the distribution of short lived species that impact the radiative balance of the planet, e.g., aerosols and ozone. Trace constituents' distribution depends on both chemistry (their sources and sinks, including those of precursors) and how they are advected and diffused by the atmospheric flow. We focus exclusively on the latter process, isolating the role of model numerics on transport from questions of emissions and atmospheric chemistry. While much recent effort has focused on improving our understanding of atmospheric chemistry, transport – largely an issue of a model's underlying numerics and resolution – still presents a challenge.

As an example, stratospheric ozone is formed primarily in the tropics (where there is most of the sunlight) and is then transported by the stratospheric circulation poleward into the midlatitudes. Hence ozone concentration, particularly in the lower stratosphere, is greater in the higher latitudes, especially in the spring hemisphere, (e.g., Dobson et al. 1926, Dobson 1956). Karpechko et al. (2013) explored factors related to the spread in ozone recovery projections by Chemistry Climate Models which participated in the CCMVal1 and CCMVal2 model intercomparison projects. They found that the spread in ozone recovery projections was most strongly associated with differences in transport (i.e., differences in how models captured the distribution of unreactive trace gases that approximate a passive tracer), more so than in differences in treatments of ozone chemistry, or models' representation of the mean climatology and variability.

The differences in ozone projections associated with biases in transport are not small. Model projections of ozone recovery (defined by the Montreal Protocol as the time when stratospheric ozone returns to 1980 levels) spans several decades. By ruling out models with demonstrably poor transport, Karpechko et al. (2013) showed that the confidence interval could be reduced by a decade.

Trace gases also play a vital role in our ability to observe the atmospheric circulation, particularly in the stratosphere (e.g., Plumb 2002). *In situ* and satellite based measurements of trace gases can be inverted to place constraints on dynamical quantities (in particular, vertical velocity) that cannot be observed directly (Linz et al. 2017). Ensuring that AGCMs provide an accurate representation of transport can thus help us both understand and quantify the observed circulation, and predict its response to external forcings.

Differences in stratospheric representation among dynamical cores has motivated several studies in the past (e.g., Yao and Jablonowski 2015, 2016). These studies, in particular, focus on investigating intermodel differences in resolved tropical processes like the QBO. In addition, while a number of short term tests have been established to test transport in numerical AGCMs (e.g. Kent et al. (2014)), to the authors knowledge there does not exist an equivalent to the Held and Suarez (1994) benchmark test to determine how the underlying numerics and resolution of an AGCM impact the climatological transport properties. This motivates us to establish a benchmark test to evaluate the impact of model numerics and resolution on the coupled stratosphere-troposphere system.

We focus on transport through the stratosphere in particular for two reasons. First, the slow time scales of stratospheric transport – on the order of months to years – provides a stiff test of model numerics. Small errors, numerical diffusion in particular, can accumulate to significant biases (Rood 1987). Second, the absence of small scale convection enables one to exclusively focus on the large scale transport by the model's numerical schemes. In the troposphere, the transport depends critically on both the resolved circulation and transport associated with parameterized moist convection (Orbe et al. 2017a,b; Wu et al. 2018).

We briefly introduce key issues in transport in Section 2, chiefly to justify our selection of the "age-of-air" as a metric to assess transport. We then propose two closely related benchmark tests for atmospheric models in Section 3. As with the HS94 intercomparison test, the goal is to remove the influence of all parameterizations on the resolved circulation and transport, allowing us to focus exclusively on the impact of a model's numerical formulation on the circulation and transport. The subtle momentum balance of the tropical stratosphere, however, proves to be a very stiff challenge for a dynamical core. In our first test, which is most similar in spirit to the original HS94 recipe, state-of-the-art numerical cores fail to represent a consistent circulation of the atmospheric circulation. Differences in the circulation make it difficult to focus in on the role of numerics on transport. This inspired a second test, where the tropical winds in the stratosphere are specified (in a manner akin to specifying the Quasi-Biennial Oscillation), which allows one to focus in on the impact of model numerics on tracer transport.

Section 4 then introduces 4 dynamical cores developed by the Geophysical Fluid Dynamics Laboratory (GFDL) and the National Center for Atmospheric Research (NCAR), which we use to establish a benchmark for the stratosphere-tropospheric circulation and transport. We are particularly interested in the behavior of the cubed sphere finite volume core developed by GFDL (Putman and Lin 2007), which recently became the core of fvGFS, the US National Weather Service forecasting system, and the cubed sphere spectral element core developed by NCAR (Lauritzen et al. 2018), which has become the main core of CESM2 for CMIP6 studies. Section 5 documents circulation and transport of the four models under our two tests. A key result is the divergent behavior between models with spectral convergence from finite volume numerics in our first, free running test. Particularly striking differences in the resolved dynamics between the models emerge as the vertical resolution is increased, which is further explored in greater detail in Section 6. We also show that once the tropical stratospheric circulation is constrained, however, the models show evidence of a more consistent representation of both the circulation and transport.

Finally, we more broadly consider the convergent behavior of the four cores in Section 7. While explicit convergence of the large scale circulation is not well defined, we find evidence of modern cores providing fundamentally different answers to stratospheric transport at high resolution. We conclude that trace gas transport still presents a challenge for model development in Section 8.

2. Stratospheric transport and the age-of-air

In the annual mean, the meridional circulation of the stratosphere draws air up from the tropical troposphere, transporting it poleward through both hemispheres before returning it back to the troposphere in the extratropics. It is known as the Brewer-Dobson Circulation, as its existence was first inferred by Brewer (1949) and Dobson (1956) based on measurements of trace gases, water vapor and ozone, respectively. The poleward transport is stronger in the winter hemisphere and the asymmetry between the summer and winter hemispheres increases with height. In the upper stratosphere and mesosphere, the circulation becomes a single cell, transporting air from the summer to winter hemisphere; the two cells appear only in the annual mean at these levels.

This overturning of mass from the tropics to extratropics, however, exists only in a Lagrangian sense. The Eulerian mean circulation of the stratosphere mirrors the multicell structure of the troposphere, e.g., with a tropical Hadley cell, but Ferrel Cells in the midlatitudes, with zonal mean equatorward flow (Murgatroyd and Singleton 1961). Outside the tropics, eddies are the primary movers of mass, effecting the poleward transport. The overturning of the stratosphere is thus characterized by the residual mean circulation (Andrews and McIntyre 1976, 1978; Boyd 1976).

The mean overturning of mass through the stratosphere is often referred to as a “diabatic circulation”, as it requires air to be radiatively warmed for it to rise through isentropic (constant potential temperature) surfaces in the tropics, and radiatively cooled to return downward through them in the higher latitudes. These diabatic processes are not the driver of the circulation, but rather a response to the mechanical forcing that draws the air poleward into both hemispheres (Haynes et al. 1991). Rossby waves, which originate from the extratropics, exert a negative torque on the stratosphere when they break, slowing the flow down so that it can move across angular momentum surfaces as it travels poleward. As described by Holton et al. (1995), these wave torques provide an “extratropical pump” that drives the circulation. The circulation is slow, with a characteristic time scales of months in the lower stratosphere, extending to years at upper levels.

In addition to driving the slow cross-isentropic advection of air, Rossby wave breaking plays a key role in transporting trace gases on faster time scales – on the order of days (McIntyre and Palmer 1983). As Rossby waves break, air is rapidly transported along isentropic surfaces. While this “isentropic mixing” leads to no net transport of mass, it can transport chemical species which exhibit a climatological gradient. Such gradients arise in response to differences in sinks and sources. Ozone, for example, is primarily produced in the tropical stratosphere, but transported poleward by isentropic mixing, and ultimately concentrated at lower levels in the extratropics by the diabatic circulation (e.g., Butchart 2014).

The transport of trace gases by an atmospheric model is sensitive to the representation of both the mean overturning circulation and isentropic mixing. An accurate representation of the mean overturning requires a model to capture Rossby wave breaking, the heat and momentum fluxes that mix potential vorticity and drive the extratropical pump. Furthermore, given the slow time scale of this overturning, quasi-vertical diffusion across isentropic surfaces by the numerical scheme can greatly accelerate the slow transport of trace gases upward in the tropics.

The isentropic mixing of trace gases by Rossby waves in atmospheric models is also sensitive to the details of the numerics. Breaking Rossby waves stretch and shear out filaments of air from the tropics and extratropics continually increasing the gradient between air masses until the point that diffusion can finally smear them out (McIntyre and Palmer 1984). As diffusive processes in the atmosphere operate on scales well below that of a model's grid (10 to 100 km), diffusion in "model world" is primarily numerical, not physical. Schemes that are more diffusive will artificially increase the impact of isentropic mixing, particularly across regions of the atmosphere that are isolated from Rossby wave breaking, chiefly the polar vortex of the winter hemisphere and the tropical stratosphere.

To study transport of tracers by both the mean overturning circulation and isentropic mixing, an idealized tracer called *age-of-air* has been used extensively in the literature (e.g., Kida 1983; Monge-Sanz et al. 2007; Engel et al. 2009; Hang et al. 2009; Linz et al. 2017), see Waugh and Hall (2002) for a comprehensive review. The age-of-air, hereafter "age", quantifies the mean time that has elapsed since air was last in contact with the surface. It is representative of how long the trace gases stay in the atmosphere once they leave the surface and, thus, provides a measure of transport timescales. For an infinitesimal, undiluted parcel of air, the age would simply be the time since it left the boundary layer. But for any finite amount of air (or for our purposes, a grid box in a numerical model), the age is the mean over the full spectrum of transit times by all the different infinitesimal parcels of air within it.

The age can be represented by a passive tracer which increases linearly in time (i.e., a parcel ages 1 second for every second of model integration), with a sink only near the surface, where it is reset to zero. In practice, however, we find it simplest to compute it using a "clock tracer", a tracer that is specified to be linearly increasing at the surface (increasing 1 unit for every second of integration, so that the concentration at the surface equals the current model time) and otherwise passively advected by the model through the free atmosphere (Waugh and Hall 2002). The age at any grid box in the model is then simply the difference between the model time and concentration at that point (see Section 3 for more details).

The key is that the clock tracer is transported by the model's advection scheme. Differences in age profiles across models provide information about differences in the representation of transport processes, both resolved (mean Lagrangian transport and isentropic mixing) and diffusive, explicit or numerical. Hall et al. (1999) computed age among 20 two and three dimensional climate models and inferred age from in-situ observations of selected trace gases in the stratosphere. The study found large differences in age between models and observations, and substantial spread across the models as well. They concluded that there were large biases in model transport schemes. Our focus on age, an integrative measure of tracer transport, complements the work of Kent et al. (2014), who analyze short-term transport of passive tracers in response to prescribed analytical wind fields.

3. Two intercomparison tests for the dynamical cores of AGCMs

We propose two closely related tests to assess the impact of model numerics on circulation and transport in dynamical cores, as summarized in Table 1. We refer to them as the "free running" (FR) and "specified tropical winds" (SP) tests. They differ only in one aspect: the treatment of the tropical stratospheric winds (bottom row of Table 1). Our tests build on the the work of HS94 and Polvani and Kushner (2002, hereafter PK02), and the new contributions of this paper are only in the bottom two rows, where we have added a clock tracer to the model and the ability to specify the tropical winds.

3.1. The Free Running (FR) intercomparison test

The key to the HS94 test is to replace all diabatic forcings (radiation, convection, etc.) with a simple temperature tendency, where the temperature is relaxed linearly toward an analytic equilibrium profile, T_{eq}, a state approximating a radiative-convective equilibrium. The diabatic forcing in the troposphere is exactly as prescribed in HS94, except for the addition of a north-south asymmetry to induce a perpetual January climatology, as detailed in PK02. The solstice climatology, as opposed to the equinox in HS94, was used because the annual cycle is critical for the stratosphere.

In order to drive a more realistic stratospheric circulation, we follow the PK02 modification of the HS94 equilibrium temperature profile above 100 hPa. This forcing imposes a fixed vertical lapse rate of -4 K km^{-1} to T_{eq} in the boreal high latitudes to approximate radiative cooling in the polar night. Away from the pole, the T_{eq} follows the US Standard Atmosphere 1976 temperature profile, with a positive lapse rate in the stratosphere. This forcing results in a cold pole in the winter (Northern) hemisphere, associated with a strong cyclonic circulation, or polar vortex. The summer (Southern) hemisphere is quiescent, with weak easterlies, as observed.

A smooth wave 2 topography of amplitude 3 km, as set up by Gerber and Polvani (2009), is added in the northern midlatitudes. The topography generates large scale stationary waves in the troposphere, providing sufficient planetary wave forcing in the stratosphere. The planetary-scale waves strengthen the wave driven circulation in the winter stratosphere and drive intermittent breakdowns of the polar vortex, Stratospheric Sudden Warmings (SSW). The -4 K km^{-1} lapse rate of T_{eq} in the winter stratosphere, paired with the 3 km topography, were found by Gerber and Polvani (2009) to produce the most realistic frequency of SSWs.

Near the surface ($p/p_{surf} \geq 0.7$), boundary layer friction is substituted with a simple Rayleigh drag to damp the low-level winds. At the model top ($p \leq 0.5hPa$), a stratospheric sponge is added to damp the winds to zero at the top boundary and to absorb any residual momentum due to upward propagating gravity waves. More details of the sponge are provided in paragraph [24] of PK02.

Age of air in the atmosphere can be computed in climate models by introducing a *clock tracer* near the surface (Hall and Plumb 1994). The clock tracer has linearly growing boundary conditions in time. In its source region near the surface, which we define to be $p \geq 700$ hPa, consistent with the height of surface friction layer, the clock tracer concentration evolves as :

$$\chi(\lambda, \phi, p, t) = t, \quad p \geq 700 \text{ hPa} \tag{1}$$

where t is the model integration time, λ is the longitude and ϕ is the latitude. Hence, the clock tracer value in the boundary region is forced to be equal to the present model time, t. Above 700 hPa, the clock tracer is passively advected with no sinks or sources, such that

$$\frac{D\chi}{Dt} = \begin{cases} 0 & p < 700 \text{ hPa} \\ 1 & p \geq 700 \text{ hPa}. \end{cases} \tag{2}$$

where D/Dt is the material derivative following a parcel. As discussed in Section 2, the age is then simply

$$\Gamma(\lambda, \phi, p, t) = \chi(\lambda, \phi, p \geq 700 \text{ hPa}, t) - \chi(\lambda, \phi, p, t) = t - \chi(\lambda, \phi, p, t). \tag{3}$$

3.2. The Specified tropical winds (SP) intercomparision tests

Ideally, we had hoped that the FR forcing could serve as our intercomparison test for dynamical cores. We found, however, that the it proved too much of a test: the dynamical cores (introduced in Section 4) failed to produce a consistent circulation, making it nearly impossible to assess the impact of numerics on transport. To motivate the need for our second test, Figure 1 shows the behavior of two different dynamical cores in response to changes in vertical resolution under the FR protocol.

The right panels of Figure 1 show the climatological circulation in two integrations of the Community Atmosphere Model - Finite Volume (CAM-FV) dynamical core (see Section 4 for further details on the models). With this dynamical core, the zonal mean winds appear insensitive to the vertical resolution. In contrast, integrations with the Geophysical Fluid Dynamics Laboratory Pseudospectral model (GFDL-PS), pictured on the left, reveal a pronounced sensitivity of the climatological circulation to changes in vertical resolution. While the differences appear only in the tropical stratosphere, the tropical winds have a pronounced impact on transport. As shown in Section 5, the age of air increases by as much as 25% in response to the formation of westerly jets in the tropical stratosphere.

While the CAM-FV model appears more "converged", i.e., insensitive to changes in the resolution, we do not believe that it is necessarily providing a more accurate representation of the circulation. The angular momentum balance in the tropical stratosphere is very sensitive, with an adjustment time scale on the order of years near the equator (Holton et al. 1995; Lauritzen et al. 2011). In Earth's atmosphere, gravity waves drive an oscillation between westerly and easterly jets over a period of approximately 28 months, the Quasi-Biennial Oscillation (QBO). Comprehensive models must be run at high vertical resolution to capture this variability (Giorgetta et al. 2002).

Figure 2(a) shows that this sensitivity of the tropical winds to model numerics and resolution extends to all four of the dynamical cores. All of the models exhibit an easterly regime at moderate vertical resolution (dashed lines), but the two dynamical cores with spectral accuracy shift to a westerly regime at higher vertical resolution (solid lines). While the CAM-FV model shown in Figure 1 appears remarkably insensitive to vertical resolution, there is evidence of a slight westerly shift at upper levels when the finite-volume model developed by GFDL is run at higher vertical resolution.

To allow a meaningful comparison of the numerical transport properties of different dynamical cores, we remove this source of uncertainty by damping the tropical stratospheric winds to a steady easterly profile in our specified tropical winds (SP) test. The damping results in consistent tropical winds across all cores (at least to 1 hPa; issues remain above due to differences in dissipation at the model top) allowing a more careful comparison of their transport properties. While we would have preferred an unconstrained test, as in the FR experiments, there is precedent for specifying the QBO winds in comprehensive atmospheric models. Except for this damping, the FR and SP tests are identical.

Damping begins at 200 hPa, and is restricted to within 15 degrees of the equator. The structure of the analytical wind profile is shown in Figure 2 (black line). The easterly profile was chosen as a neutral background state, similar to that resolved by all models at lower vertical resolution. An analytical expression and detailed description of the damping is provided in Appendix A. Figure 2(b) shows that the damping brings all models in to the same tropical circulation regime.

3.3. Experimental Setup

To test a dynamical core, we recommend an integration of 10000 days (∼27 years). The stratospheric circulation and transport evolves on time scales of months to years, such that a long integration time is needed. While the winds reach a statistical steady state relatively quickly (within a few hundred days), it takes about 6000 days for the clock tracer distribution (and hence the age) to converge to a statistically steady state at upper levels.

We use the last 8000 days of the integration to analyze dynamical quantities (e.g., the zonal winds and temperature variance), and the last 3300 days to compute the age. These long analysis periods allow us to average out internal variability, which is particular large in the winter polar vortex. As shown below, even 8000 days is insufficient to assess differences in the dynamical cores ability to represent SSWs, and therefore, much longer test integrations would need to be performed to establish significant differences between the models.

4. The Dynamical Cores

We compare four dynamical cores in this study, as summarized in Table 2. The cubed-sphere finite-volume (hereafter GFDL-FV3) and pseudospectral (hereafter GFDL-PS) cores were developed by the Geophysical Fluid Dynamics Laboratory (GFDL). The Community Atmosphere Model (CAM)-finite volume core (hereafter CAM-FV) was developed by NASA/GFDL, and the CAM spectral element core (CAM-SE) was developed at the National Center for Atmospheric Research (NCAR). We briefly highlight the key features of the four cores here, and details on their tracer advection schemes are included in Appendix B.

The GFDL-PS dycore is the oldest of the cores, and a direct descendent of the pseudospectral model analyzed in HS94. While it is no longer used by GFDL for climate prediction modeling, it is still used extensively in the idealized modeling community, e.g., serving as the dynamical core of the Isca modeling framework (Vallis et al. 2018). It uses spherical harmonics in the horizontal coupled with a finite difference discretization in the vertical. It uses the vorticity-divergence formulation of the equations and an explicitly imposed hyperdiffusion to represent sub-grid scale diffusivity.

The GFDL-FV3 dycore is GFDL's latest dynamical core, and now serves as the core of the fvGFS weather forecasting system. It employs finite volume flux-limiting schemes in the horizontal coupled with a purely Lagrangian finite volume transport scheme in the vertical. It uses a cubed-sphere grid to discretize the horizontal which implements a Gnomonic projection of the sphere on to 6 sides of a cube. Each face of the cube uses a local coordinate system native to the tile to seperately solve the equations. Such a discretization provides a quasi-uniform Cartesian-like grid which helps circumvent the pole-problem (Williamson 2007) and allows for a higher CFL timestep and provides higher scalability than the other grids. The solution from the 6 slices are then "stitched" together and interpolated onto an equispaced latitude-longitude grid for output using a 4th order mass- and energy- conserving interpolating scheme (see Putman and Lin 2007 for details).

The CAM-FV dycore is similar to the GFDL-FV3 dycore, using finite volume schemes to integrate the equations. The 2 finite volume based models majorly differ in that the CAM-FV uses a traditional latitude-longitude grid in the horizontal and uses Eulerian finite volume schemes in the vertical. The core uses the traditional latitude-longitude grid for horizontal discretizations. CAM-FV was the principle core in NCARs comprehensive climate models CESM1 and WACCM.

The CAM-SE dycore was developed to succeed CAM-FV as primary core of the CESM2 framework. It uses a continuous Galerkin spectral finite-element method to solve the primitive equations. It bears resemblance, in formulation, to both GFDL-FV3 and GFDL-PS cores as it uses a cubed-sphere horizontal discretization and a vertically Lagrangian advection scheme like GFDL-FV3 and uses hyperdiffusion as a sub-grid horizontal dissipation mechanism like GFDL-PS. In addition to using a cubed-sphere grid, identical to the GFDL-FV3 core, the CAM-SE core further uses a 4×4 spectral element, within each of its cubed-sphere cells (see Dennis et al. 2012; Lauritzen et al. 2018 for details). Similar to the FV3 core, the CAM-SE interpolates the final output to a grid appropriately chosen for the grid resolution. Hence, as shown in Table 2, results from the NE16 and NE30 integrations are interpolated onto a 256×129 and 512×257 latitude-longitude grid. This output grid resolution is not representative of the true model resolution though. See Appendix B for details.

To study how model resolution impacts transport, we compare integrations at 4 different resolutions. We first refine (double) the resolution in the horizontal or vertical, respectively, and then we refine in both the horizontal and the vertical. Table 2 summarizes the grid sizes and effective resolutions used. The model grid nomenclature (i.e., the meaning of T42 or NE16) is described in more detail in Appendix B. For brevity, we will refer to all the lower and higher horizontal resolution integrations as 2° or 1° runs respectively. Higher resolution integrations with CAM-SE model are also considered in Section 7.

5. Results: Circulation and transport benchmarks

We now compare the circulation and age-of-air distributions that result from our new intercomparison tests, considering the effects of model numerics and resolution. Circulation and age-of-air benchmarks for dynamical cores are provided, which can be used to assess newly developed dynamical cores and alternative tracer advection schemes.

5.1. Circulation benchmarks

We begin with the metrics proposed in HS94 to compare the general circulation, as shown in Figure 3. Only the results for the highest resolution runs (1° L80 runs : T85L80, C90L80, NE30L80) are shown with the exception of CAM-FV, where the F09L40 integrations are used. The 80 layer integrations in this core were anomalous, particularly for the specified tropical wind test, and would skew the overall result. The figure shows the zonal mean zonal winds, eddy temperature variance, and the zonal spectra of eddy kinetic energy. The same metrics at select pressure levels are shown in Figure 4, which allows a comparison between the individual models. In all line plots throughout the section, the dashed curves represent the free running (FR) integrations, and the solid curves represent the runs with specified (SP) tropical winds.

The contours in Figure 3(a) show the ensemble mean zonal mean wind patterns from FR runs, and the red shading shows the standard deviation across the 4 run ensemble. Outside the tropical stratosphere – differences that we highlighted already in Section 3 – all integrations capture nearly the same large-scale circulation. The intermodel spread is largest in the extratropical stratosphere where the strength of the winter polar vortex (the westerly jet with peak wind speeds $\approx 50\ m/s$ near $70^{\circ}N$ and 3hPa) varies between models. This can be seen more clearly in Figure 4(b). There are also subtle differences in the extratropical troposphere, namely in the position of the midlatitude westerlies, which do barely appear at this contour interval, but can be seem more clearly in Figure 4(a). The subtropical and extratropical jets are merged in this climatology, more so than in the original HS94 test, but there are hints of greater separation between them in the finite-volume based cores, particularly in the summer hemisphere.

In the ensemble mean, the tropical stratosphere exhibits easterly winds, but the variance is large. Figure 4(b) shows the split between the spectral-based cores, which develop westerly jets, and the finite volume cores, which maintain easterlies. This split in behavior is eliminated in the specified tropical wind (SP) integrations shown in Figure 3(b). Variance between the tropical stratospheric winds in this test is removed by construction, although there are still differences at upper levels, particularly near the model top (not shown). Outside the tropical stratosphere, the winds in the SP and FR integrations are pretty comparable, as is the spread between models. There is more spread in the strength of the stratospheric polar vortex; in GFDL-FV3, specifying the tropical winds leads to a weaker vortex, while in CAM-FV, the vortex increases in strength.

Figure 3(c) shows the ensemble mean eddy temperature variance in the troposphere for the SP runs; the results are nearly identical in the FR integrations, and so not shown. Our choice of a perpetual northern winter climatology endows a hemispherical asymmetry in the eddy temperature variance structure not observed in the HS94 test, the variance being much larger in the winter hemisphere. The variance is maxiumum near the surface and decreases higher up in the troposphere up to roughly 400hPa. Above that, variance in the upper troposphere slightly increases, associated with the midlatitude jet variability and baroclinic eddy heat exchange between stratosphere and troposphere.

Figure 4(c,d) shows the temperature variance in the individual models for select levels. At 300 hPa, which captures the upper troposphere maximum, variability is weaker in both of the finite-volume based cores relative to the spectral based cores. This suggests that finite volume cores are more strongly damped; although dominated by synoptic scales, this variability increases with resolution (not shown). Figure 3(c) was truncated at 100 hPa, as the variability in the polar vortex is very large, exceeding 150 K^2 at 10 hPa. Differences in the variance at 30 hPa, shown in Figure 4(d), are characteristic of differences at other levels. There is notably less temperature variability of the vortex in the GFDL-FV3 integrations.

Figure 3(d) shows the vertically integrated zonal spectra for the eddy kinetic energy (EKE) for the SP runs. The concentration of EKE at wavenumber-2 is associated with stationary waves induced by the topography, and the asymmetry between the hemispheres (not present in HS94) is due to the winter time climatology added by Polvani and Kushner (2002). We do not show this metric for our FR runs, as the vertically integrated spectrum for the ensemble is nearly identical for both the FR and SP tests.

5.2. Coupled stratosphere-troposphere variability benchmark: Sudden Stratospheric Warming events

The wave-2 topography was developed by Gerber and Polvani (2009) to increase the planetary scale wave forcing on the stratosphere, and to obtain a realistic Sudden Stratospheric Warming (SSW) events. A SSW is identified by an abrupt reversal of the stratospheric polar vortex, specifically the winds at 60° N and 10 hPa, following World Meterological Organization (WMO) convention (Mcinturff 1978). Figure 6(a) illustrates a time series of the polar vortex winds in CAM-SE, a thousand day period during which there were three SSWs (marked by vertical dashed lines), in addition to a minor warming at the start of the period, when the vortex weakened, but did not completely reverse. The break down of the vortex is associated with a dramatic warming of the polar stratosphere, which occurs approximately every other year in the boreal hemisphere (e.g., Butler et al. 2017). SSWs are a source of surface weather predictability on subseasonal to seasonal time scales, as a poleward shift of the tropospheric jets follows the breakdown of the vortex (e.g., Kidston et al. 2015).

Climate prediction models have struggled to properly capture the correct frequency of SSWs (Charlton-Perez et al. 2013). In general, reasonably fine vertical resolution of the stratosphere and a model lid above the stratopause is needed (albeit not sufficient) to simulate them. Figure 5 suggests, however, that the choice of model numerics may also play an important role. Overall, most of the dynamical cores simulate between 2 and 3 SSWs per thousand days of integration. In observations, SSWs are observed every other year, but recall that we simulate a perpetual winter state. While the GFDL-PS and CAM-SE models' SSW statistics appear insensitive to changes in horizontal or vertical resolution, the two finite volume based cores appear sensitive to horizontal resolution.

The CAM-FV core stands out at low (2°) horizontal resolution, simulating very few events (if at all). When run at 1° resolution, however, it's statistics become consistent with the other models (with the exception of the L80 run with specified tropical winds, which appeared anomalous in many respects). In contrast, the GFDL-FV3 model appears to simulate only half as many warmings than the other dynamical cores when run at higher horizontal resolution. The fact that SSWs are rare events makes it difficult to establish the statistical significance of this difference. Even with 8,000 days of integration time; the 2-sigma error bounds (not shown) still overlap.

Limitations on available computational resources precluded us from further investigating the SSW statistics of the 1° resolution integrations. We did probe the 2° L80 runs in more detail, however, running three of the models for 100 years each. This provided a factor of 4 increase in sampling, and hence narrowed the uncertainty of the mean frequency by a factor of 2, as shown at the far right of Figure 5. Based on these results, we believe that 100 years of integration time would be sufficient to clarify whether GFDL-FV3 behaves significantly different at 1° resolution.

It might come as a surprise that the SSW statistics are most sensitive to horizontal resolution; this is due to the fact that even the 40 level integrations exhibit a well resolved stratosphere compared to a typical CMIP5 model. Not only is 40 layers fairly high vertical resolution, the simplicity of the HS94 forcings does not require fine resolution of the troposphere, allowing us to place 17 levels between 100 and 1 hPa. The dependence on horizontal resolution is not easy to explain. The 2° CAM-FV model simulates an anomalously strong stratospheric polar vortex, which may inhibit planetary wave propagation. But for the 1° GFDL-FV3 integrations, the vortex is actually anomalously weak, as seen in Figure 4(d). This is associated with greater wave forcing on average (and a correspondingly larger meridional overturning circulation), but the wave forcing is evidently more uniform in time.

When the tropical winds are specified, a slight (albeit consistent) increase in the average sudden warming frequency is observed for all cores at all resolutions (again with an exception of the CAM-FV core). This increase can be explained by the (Holton and Tan 1980) effect. Tropical easterlies inhibit the equatorial propagation of planetary scale Rossby waves, trapping them in the high latitudes, leading to a more disturbed polar vortex (and hence, more SSWs). In contrast, a westerly jet in the tropics shifts the critical lines for planetary waves equatorward, allowing them to percolate into the tropics (and even the summer hemisphere). This reduces the planetary wave activity in the winter midlatitudes, resulting in fewer SSW events (Shuckburgh et al. 2001; Garfinkel et al. 2012).

The polar stratospheric vortex is associated with sharp meridional gradients of potential vorticity, which inhibit exchange of air between the polar and midlatitude stratosphere, allowing very old air to accumulate in the polar vortex (McIntyre and Palmer 1983; Sobel et al. 1997). The breakdown of the polar vortex during an SSW destroys this barrier to mixing, homogenizing air between the midlatitude surf zone and the pole. A signature of this can be observed in the age of air, as illustrated in Figure 6(b,c). Overall, the concentration of the clock tracer averaged over the polar cap (panel b) grows linearly in time, indicating that the age of air in the polar vortex has converged (panel c). This growth, however, is highly episodic, accomplished almost exclusively during SSWs, when young air (with a clock tracer concentration closer to the surface value) is mixed into the vortex. The tracer concentration spikes by several hundred "days" over the course of a month, associated with a drop in the age of air of over a year (panel c).

After a warming event, the polar vortex re-establishes itself, isolating the polar air again, such that the concentration of the clock tracer flat lines. It actually decreases in some cases, particularly between days 9400 to 9600, as very old air descends slowly into the vortex. It is these large variations in the age-of-air in the polar vortex associated with SSWs that required a long averaging period (3000 days) to obtain a reliable estimate of the mean age. For most other regions, the age-of-air can be reliably estimated with a much shorter averaging window.

5.3. Transport Benchmark: Age-of-air

Figure 7 shows the ensemble mean age-of-air in the stratosphere, taken across the highest resolution runs as in Figure 3. The left panel shows results based on the FR test and the right, the SP test; in both, solid contours show the age-of-air and the color shading, the fractional deviation in the age across the ensemble mean. The most conspicuous difference between the two tests is the fractional age deviation: the FR integrations vary by a factor of up to 25% in the tropical stratosphere and winter midlatitudes, corresponding to a spread of up to 2 years in the age of air. The deviation for the SP test are substantially lower throughout the stratosphere, (just 6% in the tropics). A closer look also reveals significant differences in mean age between the 2 ensembles. For instance, the ensemble mean age at 45N and 10hPa is over 10% higher in the FR runs compared to the SP integrations.

Before investigating these differences, we first highlight the large-scale structure of the transport profiles. The overall structure of the mean age in the models is consistent with observationally based estimates of the mean age of the atmosphere, but with an old bias. The age in the upper stratosphere exceeds 9 years in the models, compared to 6 years in observations (Waugh and Hall 2002, their Figure 7). This is partly due to the fact that tropopause is artificially low in the PK02 stratosphere (such that the air at 100 hPa has already spent considerable time in the models' stratosphere), but also indicates that overall circulation is weak compared to that of our atmosphere. Gravity wave driving plays an important role in the meridional overturning of the stratosphere, particularly at upper levels, but is only crudely approximated in these integrations by sponge layer above 0.5 hPa. The slower circulation, however, helps emphasize differences in numerical schemes, as small errors accumulate over time.

In the vertical, the stratosphere always becomes older with height, even in the midlatitudes where the mean transport of mass is downward. While the mean overturning tends to bring air upward in the tropics and downward in the mid and high latitudes, isentropic mixing between the upward and downward branches of the cells ensures that the oldest air is found aloft. The horizontal structure of the age distribution, which can be more clearly seen in Figure 8, reveals three distinct regions: (i) the tropical pipe, where comparatively young air is brought up into the stratosphere, (ii) the midlatitude surf zone, where isentropic mixing homogenizes the age (see most clearly at 30 hPa in Figure 8b), and (iii) the polar vortex, where on average the oldest air is found. The three regions are seperated by sharp gradients in age, associated with barriers to mixing. The tropical barrier, c. 15-20°N is associated with tropical limit of planetary scale wave breaking, while the polar barrier at 60°N is associated with strong potential vorticity gradients surrounding the strong polar vortex.

We have focused on the northern (winter) hemisphere, as perpetual summer in the southern hemisphere makes it rather artificial. Easterly flow in the summer hemisphere inhibits Rossby wave propagation, limiting both isentropic mixing and the meridional overturning circulation. Overturning between the tropics and high latitudes drops off abruptly above 50 hPa, virtually dissappearing at 10 hPa, above which there is just a single cell, transporting mass from the summer to winter hemisphere. This leads to very weak net advective transport in the region and diffusion plays a primary role in effecting transport here.

Returning to the difference between FR and SP results, the key factor contributing to the large age variance in the FR integrations is a split in the vertical structure of age in the spectral-based models vs. the finite volume-based models. This is immediately evident in Figure 8, which shows the age-of-air among different models at 3 different pressure levels in the stratosphere, chosen to represent the lower, middle and upper stratosphere.

At 150 hPa, the age is very similar among all models, independent of the model numerics (colors) or treatment of tropical winds (dashed vs. solid). The strong gradient between the tropics and higher latitudes reflects the contrast between fresh air brought up into the stratosphere in the tropics and the comparatively old air returning down to the troposphere at higher latitudes. The gross age gradient is a function of the diabatic circulation strength alone, independent of isentropic mixing, and reflects the consistency of the meridional overturning across the models.

Higher up, at 30 hPa and 10 hPa, however, a key differences between the dynamical cores emerges in the FR integrations (and, but to a much smaller extent, in the SP integrations). In the free running integrations, the spectral-based cores (GFDL-PS and CAM-SE) exhibit an older age-of-air all the way from the subtropics of the summer hemisphere to the winter pole (dashed blue and green). The air in the midlatitudes is as much as 2.5 years older (i.e., around 25% older, at lower levels) when compared to the finite-volume based cores (dashed red and orange). The spread between the finite volume and spectral models grows slightly with height, but less so than the mean age, so that the relative (fractional) variance decreases with height (Figure 7).

Specifying the tropical winds has little effect on the finite volume models, but leads to a significant decrease in the age in the spectral models. The age in the winter midlatitudes drops by as much as 2 years. This explains both the drop in mean age and the dramatic drop in the spread between the FR and SP ensembles. We were struck by the consistency between the two models with finite volume numerics and for the two model with spectral accuracy. Perhaps this should not come as a surprise for the finite-volume models, which differ primarily in their grid (lat-lon vs. cubed sphere), although they have different treatment of vertical advection. The two "spectral" models are rather different, particularly in the treatment of tracer transport; GFDL-PS is a very old core based on decomposition of the flow into the spherical harmonics, while CAM-SE is a very modern core based on a cubed-sphere grid. We therefore speculate that differences stem from their representation of the resolved dynamics, where the use of explicit hyperdiffusion makes them less diffusive on intermediate scales.

The region with peak ensemble age variance in the FR experiment coincides with spread in winds (Figure 3, top left plot). As seen in Figure 2, the winds abruptly shift abruptly toward a westerly jet at about 80 hPa in the two spectral models. The age in the tropics diverges between the models at this level, with the spectral models remaining consistently older at all levels above. The difference in age spreads all the way to the pole, however, even though the winds are similar outside the tropics. As seen in Figure 8, the difference between the finite volume and spectral models increases a bit with height, but to first approximation, the difference could be attributed to a uniform (in latitude) jump in age at 80 hPa, which remains roughly constant at all levels above. This abrupt change in the vertical gradient of age suggests a difference in mixing, which we investigate in Section 6.

Even when the the tropical winds are specified, there is a split between the age in the spectral and finite volume models. The spectral models consistently show older age throughout the winter stratosphere, by as much as 0.75 to 1 years in the winter extratropics. Moreover, the reduction in the ensemble variance is only seen in the winter hemisphere and summer tropics, as the age-of-air in the extratropical summer hemisphere remains largely unchanged. These differences between the age in the spectral and finite volume models, and it's dependence on resolution, is investigated in more detail in Section 7.

6. Sensitivity to vertical resolution

Transport in the stratosphere is affected by both the Lagrangian mean overturning of mass and by the mixing of tracers along isentropes (layers of constant entropy, or potential temperature). An increase in the meridional overturning will enhance the rate at which the stratospheric trace gases are flushed out of the region, thus reducing the age-of-air, all other things being equal. As discussed by Linz et al. (2016), the meridional overturning can be directly linked to the horizontal gradient of age. An increase in the isentropic mixing, on the other hand, results in an overall increase in the age in both the tropics and extratropics, as seen in the leaky pipe model (Neu and Plumb 1999).

6.1. Differences in the Diabatic Circulation

To quantify Lagrangian overturning circulation, we compare the isentropic mass streamfunction. The streamfunction represents the mean meridional mass transport in the stratosphere and provides a quantitative estimate of the Brewer-Dobson circulation (the overturning circulation of the stratosphere), to a good approximation (Townsend and Johnson 1985). As in Pauluis et al. (2009), the Eulerian streamfunction in isentropic coordinates can be expressed as an integral of the meridional mass flux as :

$$\psi(\phi,\theta_0) = \frac{Rcos\phi}{g}\frac{1}{T}\int_0^T\int_0^{2\pi}\int_0^{p_s} vH[\theta(\lambda,\phi,p,t)-\theta_0]dpd\lambda dt \quad (4)$$

where ψ is the mass streamfunction, $\lambda \in [0, 2\pi]$ is the longitude, $\phi \in [-\pi/2, \pi/2]$ is the latitude, θ is the potential temperature at a given position, R is the radius of the earth and g is the acceleration due to gravity.

Figure 9(a) shows the ensemble mean diabatic circulation (in black) among the highest resolution FR runs (with an exception of CAM-FV as before). The ensemble mean diabatic circulation strength is shown using solid black contours and the colors show the standard deviation in the circulation strength across the ensemble. Subplot (a) shows the ensemble mean diabatic circulation for the FR integrations.

The obtained circulation profile conforms to the observed mean meridional stratospheric overturning circulation. In particular, it suggests that in a mean Lagrangian sense, the air moves upwards in the tropics and downwards in the higher latitudes, while gradually moving polewards. A perpetual winter results in a much stronger mass transport in the winter hemisphere and higher up in the stratosphere the tropical diabatic upwelling intrudes all the way into the summer hemisphere.

Similar to the trends in zonal winds and ensemble mean age-of-air (Figure 3 and 7 respectively), the models exhibit strong variations in the diabatic circulation strength in the tropics (red shading, Figure 9). Between 450K and 600K, the circulation strength varies by as much as 15%. In contrast to the zonal wind variance, these differences pervade all the way into the winter midlatitudes. Subplots (b)-(d) show the variations among different cores at 3 different levels in the stratosphere. The models tend to agree on the circulation structure and strength at all levels, except at 500K, where they disagree in the tropics and winter midlatitudes. Interestingly, the GFDL-PS and the CAM-SE cores, which resolve tropical westerlies, tend to have a much weaker diabatic circulation in the region, as compared to the finite-volume cores. This difference between the spectral-based cores and the finite-volume cores results in the high variance obtained in Figure 9(a).

These variations in mass transport are eliminated once the tropical winds are specified, as seen in Figure 9(e) which shows the same metrics as in Figure 9(a), but for the SP integrations. The models no longer disagree on the diabatic circulation strength in the tropics. A closer look at 3 different isentropic levels (Figure 9(f)-(h)) in the stratosphere reiterates this point all the cores agree on the circulation strength at all levels. This confirms that differences in mass transport were indeed due to differences in the resolved tropical winds among the cores and establishes the sensitivity of diabatic mass transport to the tropical stratospheric winds.

Tropical westerlies in the spectral cores induce a secondary mean meridional circulation in the tropics. This is in agreement to past studies (e.g. Plumb and Bell 1982) where the easterly and westerly phases of the QBO were found to induce opposite secondary circulation in the tropics. While the easterly phase induces an clockwise circulation in the tropics, the westerly phase induces an anti-clockwise circulation.

It should be noted that even specifying the tropical winds does not eliminate the inter-model differences altogether. As is seen in Figure 9(a) and (e), while specifying the tropical winds eliminates the uncertainties in circulation strength in the tropics, it also leads to a slight increase in circulation spread at higher latitudes in the lower stratosphere (around 400K). This spread is seen more clearly in Figure 9(h), where the models tend to have slightly different circulation at 400K in the 45-75N region. Even as the spread reduces as we go higher up in the stratosphere, Figure 9(g) indicates that the GFDL-FV3 core maintains a slightly stronger circulation strength than the other cores at 500K.

6.2. Tropical westerlies and isentropic mixing

Jets associated with the QBO also impact isentropic mixing between the tropics and midlatitudes. Enhanced mixing between the tropics and the extratropics is observed during the westerly QBO phase, as the westerly winds allow midlatitude Rossby-waves to penetrate deeper into the tropics. Theoretical models of stratospheric transport in turn demonstrate that enhanced mixing will increase the age-of-air throughout the stratosphere (Neu and Plumb 1999). The enhanced mass flux transports (mixes) older midlatitude air into the tropics, increasing the age in the tropics. Naively, one might expect the age of midlatitude air to deccrease (as younger tropical air is brought in), but this affect is counteracted by an increase in the age of the air circulating to the extratopics from the tropics (see Neu and Plumb 1999; Linz et al. 2016, for more detailed explanation).

To establish that the tropical jets drive this mixing effect, we perform a second specified wind profile experiment with the GFDL-FV3 core. This allows us to compare the SP integration, with easterly winds throughout the tropical stratosphere, to a second integration, where the same model is forced to a westerly profile. The westerly profile imposed on GFDL-FV3 is taken from the 2° L80 (NE30L80) FR integration with the CAM-SE model (Figure 2, in solid green).

Figure 10 compares the diabatic circulation (light curves) and age-of-air (dark curves) between the GFDL-FV3 and CAM-SE runs at 3 different isentropic levels in the stratosphere. At 380K in the lower stratosphere, Figure 10(c), the tropical winds have virtually no affect on the age-of-air and mass transport. While there are some differences in the age-of-air between 15N and 40N at 380K, these seem to be unrelated to the tropical winds themselves, rather reflecting diffrences in synoptic-scale mixing between the two cores.

The situation changes as we proceed higher up in the stratosphere. At 550K ($\sim$ 50hPa), there is a split between integrations with tropical westerlies (dashed) and tropical easterlies (solid), Figure 10(b). The diabatic circulation is reduced in both GFDL-FV3 and CAM-SE integrations with tropical westerlies. The change in mass transport is accompanied by a change in the age: it is the tropical winds that control differences in transport, not the differences in the numerics of the cores.

At 800K ($\sim$ 10hPa), Figure 10(c), the diabatic mass transport is similar across all the integrations, consistent with the fact that at this level, the climatological winds are similar among the 4 runs. The differences in age-of-air, however, continue to increase. This is a signature of a change in mixing. The increase in isentropic mixing across the barrier between the tropics and midlatitudes, facilitated by the westerly jets, makes the air older throughout the whole stratosphere *above it* (Neu and Plumb 1999). At 800K, the GFDL-FV3 and the CAM-SE westerly runs exhibit nearly identical age profiles, both nearly two years older than their easterly counterparts.

The findings suggest the critical importance of tropical momentum balance in stratosphere-resolving models. In comprehensive climate models, biases in representation of tropical stratosphere would lead to significant biases in transport of active trace species including ozone and water vapor. While tuning a gravity wave parameterization to force the QBO can help eliminate these biases, it's unlikely that the parameterization is providing an accurate representation of gravity wave activity; rather it's being used to correct for other biases in the model numerics.

7. Numerical impacts on transport

Specifying the tropical winds in the SP test brings all the models into the same flow regime, independent of resolution. As observed in Figures 7(b) and 8, however, differences persist in their representation of age. In particular, we still observe a split between the cores with spectral vs. finite volume numerics, although the difference is not nearly as large as in the FR runs. It is tempting to attribute this to fundamental difference between finite volume and spectral numerical schemes. Apart from their spectral convergence, however, the numerical schemes in the GFDL-PS and CAM-SE differ in almost all respects. Likewise, the two FV cores have been built on different grids, with important differences in their treatment of vertical advection. To further probe these differences, we document differences in the age distribution as a function of model resolution.

7.1. The age distribution as a function of resolution

Figure 11 compares the age-of-air across all SP integrations at 2 levels chosen to characterize the structure of the lower and upper stratosphere. We first focus on the winter (northern) hemisphere. For the CAM-FV core (red), the age throughout much of the winter midlatitudes and poles is very sensitive to both vertical and horizontal resolution. The integrations with higher vertical resolution (F19L80 and F09L80) exhibit notably older air in the polar region, indicating a more isolated polar vortex. This is consistent with the relatively low frequency of sudden warming events in the core, as seen in Figure 5. As noted earlier, we are concerned that the SP runs with this model may not accurately reflect the performance of the core. Following HS94, we consider the small scale damping a part of the numerical method, and used the default settings for all cores. For CAM-FV, this entailed second-order divergence damping. Lauritzen et al. (2012) found that CAM-FV simulates inordinately high polar vortex winds at high horizontal resolution (1° and 0.5°), unless fourth-order divergence damping is used.

The age simulated by the other models is more robust to changes in resolution, particularly in the more modern cores (GFDL-FV3 and CAM-SE). In the GFDL-PS core, transport is sensitive to vertical resolution. The age throughout the winter hemisphere is nearly identical in the two low vertical resolution (T42L40 and T85L40; dashed blue lines in Figure 11c,d). At higher vertical resolution, the age increases, particularly in the T85L80 integration (dark blue), where age within the polar vortex increases by approximately 1 year. The increase in age with vertical resolution is consistent with a reduction of vertical diffusion, an expected impact of increasing the vertical resolution.

The age increases with vertical resolution in the GFDL-FV3 and CAM-SE cores as well, albeit to a lesser extent. Transport by the GFDL-FV3 model is remarkably insensitive to the horizontal resolution; the age profiles for the two 40 layer and 80 level models lie atop another, respectively. The split between L40 and L80 integrations is not simply an issue of diffusion, however, as the age in the tropics is consistent across all the GFDL-FV3 integrations. Rather, there is an increase in the gradient between the tropics and extratropics with resolution, which indicates a decrease in the overturning circulation (Linz et al. 2016). A weaker overturning circulation implies a reduction in wave driving, consistent with the reduction of the frequency of SSWs in Figure 5. We therefore speculate that the key difference is the sensitivity of the model's resolved circulation, not the sensivity of the tracer transport scheme.

As with GFDL-FV3, CAM-SE's respresentation of the age is fairly insensitive to changes in resolution. The age profile is nearly identical for all the runs (Figure 11(a,b) green curves) with a minor exception of the highest resolution integration, NE30L80, which develops a slightly older age (about 0.2 years) throughout the winter hemisphere. For this core alone, we were able to perform two additional integrations, doubling the horizontal resolution (NE60L80, half degree resolution) and doubling the vertical resolution (NE30L160, where the horizontal resolution is kept at 1 degree). The age in both of these integrations lines up on top of the NE30L80 integration, indicating that the age distribution has converged.

The summer hemisphere in our idealized forcing is perpetually quiescent, exhibiting weak isentropic mixing and diabatic overturning (Figure 9). It therefore provides a stiff test on the model numerics, as diffusion becomes comparatively more important. The asymmetry between the hemispheres increases with height, leading to greater variance in age at 30hPa compared to 65hPa. The GFDL-PS model at lower vertical resolution stands out in particular. For all models, however, the age increases and becomes more consistent in the high vertical resolution integrations, suggesting the importance of limiting spurious vertical diffusion by the numerical scheme.

7.2. *Numerical impacts on transport*

Our results highlight the importance of employing high vertical resolution to accurately model stratospheric transport. Accurate simulation of the formation and recovery of the ozone depends not only on a proper treatment of the chemistry, but also an accurate representation of the transport (Karpechko et al. 2013). We also find fundamental differences between models employing finite volume vs. spectral schemes. The age distribution in both the GFDL-FV3 and CAM-SE models show evidence of convergence, but the models are converging to different climatologies. For example, the age in the midlatitude surf zone at 30hPa and 45N is 6 years in GFDL-FV3 is, but 7 years in CAM-SE. This difference exceeds the internal spread between the integrations of each model with varying resolution, respectively.

There are two key structural differences between the models. First, the age within the tropics is more homogeneous in the spectral cores, with little asymmetry between the hemispheres. In the two finite volume cores, air is younger right at the equator, with a sharper gradient in the summer hemisphere. This indicates differences in isentropic mixing within the tropics. Second, the spectral cores tend to age more strongly in height; the gulf between CAM-SE and GFDL-FV3 widens at 30 hPa relative to 65 hPa. This difference in the vertical gradient indicates a higher rate of isentropic mixing between the tropics and extratropics in CAM-SE. Both of these differences therefore point to the representation of mixing and diffusion along isentropic surfaces. The fact that the models exhibit a similar mean overturning circulation indicates that they agree on the mixing of potential vorticity. Hence this may be exclusively an issue of how trace gas transport is represented by the numerical scheme.

Differences within the tropics could be related to the conservation of Axial Angular Momentum (AAM). Past studies have focused on notable differences in angular momentum conservation between spectral and finite volume based cores. In particular, Lauritzen et al. (2014) finds that the CAM-SE core, like the GFDL-PS core, conserves AAM very well. Moreover, the models capability to conserve AAM is not significantly affected by their vertical advection schemes. Such is not the case for the two finite volume based cores. Studies focused on superrotating extra-terrestrial atmospheres have found the spectral cores to represent the super-rotating jet more accurately than the finite volume cores, indicating that the finite volume cores might have a problem with angular momentum conservation. It is unclear how angular momentum conservation would affect tracer transport, but it may explain the differences in the FR integrations. Conservation of axial angular momentum would support the development of westerly jets (super-rotation) in the tropics.

Numerical dissipation may be another important source of differences between the cores. Finite volume cores employ flux-limiters to preserve stability and monotonicity in higher-order dispersive schemes, which can render them highly diffusive. The spectral cores tend to be less dissipative, particularly for small scale waves. The effects of diffusion are the strongest on small scales. Yao and Jablonowski (2015, their figure 12) found that the power spectrum from the CAM-FV cores was much weaker than that of the spectral cores. While they do not conclude that the weaker wave activity in CAM-FV is necessarily wrong, these differences in the representation of waves can potentially be a crucial factor in the tropical stratosphere.

Dissipation on small scales impacts the effective resolution of a model. The "effective order of accuracy" allows one to compare the numerical truncation errors as a function of resolution. The schemes employed by the GFDL-PS and CAM-SE cores are spectrally accurate, i.e., exhibiting exponentially increasing accuracy with resolution, while the finite volume schemes in GFDL-FV3 and CAM-FV are second- or third- order accurate. We emphasize that this payoff in accuracy is achieved only in the limit of infinite resolution. For the resolutions considered in this study, it's unclear which model is most accurate.

Differences in the representation of small scale gravity waves, however, could help explain differences in the FR integrations in particular, as their momentum fluxes play a key role in sustaining westerly jets in the tropical stratosphere. Too much noise on small scales, however, could lead to artificially strong mixing along isentropes, explaining the homegenization of age within the tropics in the spectral models compared to the finite volume cores. Small scale noise could also increase mixing between the tropical pipe and midlatitude surf zone, and hence the increase in age with height observed in spectral models. This claim is speculative, and the subject of current research.

8. Conclusions

We have proposed two tests to assess the climatological representation of the stratosphere-troposphere system in primitive equation models on the sphere, the dynamical cores of atmospheric general circulation models (AGCMs). They are a natural extension of the Held and Suarez (1994) intercomparison test, updated to probe the representation of stratosphere-troposphere coupling and tracer transport. An active stratosphere was established by using the equilibrium temperature profile developed by Polvani and Kushner (2002) and the addition of surface topography (Gerber and Polvani 2009) to promote planetary scale stationary waves. Tracer transport was assessed by including a clock tracer to allow computation of the age-of-air (e.g., Waugh and Hall 2002). Age-of-air is a measure of transport timescales in the atmosphere, allowing one to diagnose the combined influence of the mean Lagrangian transport and isentropic mixing, as reviewed in Section 2.

To establish a benchmark for comparison, four different dynamical cores developed by the Geophysical Fluid Dynamics Laboratory (GFDL) and National Center for Atmospheric Research (NCAR) were assessed. This allowed us to compare pseudospectral, spectral element, and finite volume numerical schemes employing latitude-longitude and cubed sphere grids. The robustness of the circulation and transport to changes in model resolution was probed by integrating the models at 4 different resolutions, refining the grids in the horizontal and vertical directions, individually and in concert. The dynamical cores explored include GFDL-FV3, the finite volume cubed sphere core of the National Weather Service forecasting model and CAM-SE, the spectral element model that will be the core of CESM2.

The first test, referred to as FR for "free running", is closest in spirit to the Held and Suarez (1994) test, differing only in the revised temperature equilibrium profile and orography. The four models fail, however, to develop a consistent representation of the large scale circulation, particularly at high vertical resolution. We observed a split between the two models with spectral accuracy (GFDL-PS and CAM-SE), which developed westerly jets at high vertical resolution, compared to the two cores with finite volume numerics (GFDL-FV3 and CAM-FV), which tended to keep the easterly jets observed in all models at lower vertical resolution. It is well known that the momentum balance of the tropical stratosphere is subtle; this is the region of the atmosphere where the Quasi-Biennial Oscillation (QBO) is observed, a slow oscillation between westerly and easterly jets over a period of approximately 28 months.

While the circulation outside the tropical stratosphere was largely unaffected by these differences, tracer transport throughout the entire stratosphere was profoundly affected. Westerly jets induce enhanced mixing along isentropic surfaces, as discussed in Section 6, leading to significant aging of the air throughout the winter stratosphere, by a factor of 25% or more. The large differences induced by the circulation made it nearly impossible to assess the importance of model numerics in tracer transport.

This motivated our second test, which we refer to as SP for "specified tropical winds", where the zonal winds in the tropical stratosphere are constrained by a Rayleigh drag to a prescribed easterly profile. This prescription of the tropical winds is justified in part by the fact that our idealized forcing does not provide a realistic representation of gravity wave sources in the troposphere, and the resolution of the models is insufficient to accurately model the propagation and breaking (and so momentum deposition) of gravity waves in the free atmosphere. Prescribing the winds is somewhat akin to specifying, or tuning the gravity wave parameterization to capture, the QBO. The circulation outside of the tropics is largely unaffected by this additional momentum forcing, although we do observe an increased frequency of Sudden Stratospheric Warming (SSW) events across all the models, consistent with the observations of Holton and Tan (1980), who found that tropical easterlies inhibit the propagation of planetary waves into the tropics, thereby steering them into the higher latitudes, inducing a weakening of the polar vortex.

With specified tropical winds, the model's representation of the large scale circulation is fairly consistent across the four cores. This reassuring result provides a modern update to the comparison between a pseudo-spectral and finite difference core by Held and Suarez (1994) 25 years ago. Despite differences in underlying numerics, domain discretization, and resolution, the models exhibit overall agreement on the structure of the large-scale circulation and eddy statistics, as documented in Section 5. Key differences remain, however, suggesting that model numerics still matter for the representation of the stratosphere-troposphere system. In the troposphere, the position of the extratropical jets varies by approximately a degree of latitude; while this difference is small relative to the grid spacing, it's on the order of projected changes in the jet stream (Barnes and Polvani 2013) and consistent with biases in CMIP5 models (Wenzel et al. 2016). In the stratosphere, there are notable differences in the strength of the polar vortex, the extent of the meridional overturning circulation into the polar region, and the frequency of SSWs (although the statistical significance of this final result could not be established, even with 8,000 days of integration).

We chose the age-of-air as a measure of models' transport because it captures the integrated effect of the mean Lagrangian circulation of mass and the mixing of trace gases along isentropic surfaces. As shown by Waugh and Hall (2002), it also provides information on the transport of actual trace gases by the atmospheric circulation. Figure 7 establishes a benchmark for the age-of-air under the (Polvani and Kushner 2002) forcing with topography. While differences remain, the four models provide an overall consistent estimate of the age provided the tropical winds are constrained.

As illustrated in the analytic "leaky pipe" model of Neu and Plumb (1999), and applied to full three-dimensional models by Linz et al. (2016), horizontal gradients in the age can be related to the mean overturning circulation. As explored in Section 6, we find some spread in the representation of the mean overturning circulation, particularly in the high latitudes. The spread in transport in the FR runs, however, is dominated by differences in isentropic mixing. While the vertical gradient of age depends in part on the mean overturning, it is set primarily by isentropic mixing. By comparing GFDL-FV3 and CAM-SE models with easterly and westerly jets, we found that changes in isentropic mixing driven by differences in the zonal wind structure dominate the changes in age in the FR integrations.

Even when the tropical winds are constrained, however, structural differences between the cores with spectral accuracy and finite volume numerics persist. In particular, the age-of-air in the GFDL-FV3 and CAM-SE cores appears very stable to changes in the vertical and horizontal resolution, but the two models converge to different age distributions, as highlighted in Figure 11. The horizontal gradients in age are fairly consistent, in keeping with the similarity in the mean overturning circulation, but CAM-SE exhibits a steeper gradient of age in the vertical, suggesting a higher rate of isentropic mixing. In addition, the finite volume models indicate stronger gradients of age within the tropics, with more asymmetry between the summer and winter hemispheres. This suggests that transport remains sensitive to model numerics, and future work will seek to assess the processes responsible for these changes.

This study has assessed the state of the general circulation and tracer transport in dynamical cores. Although the finite volume cores provide a more consistent circulation across resolution in the FR test, we do not believe this necessarily indicates a better representation of the dynamics. We speculate that the spectral cores are potentially providing a more accurate representation of momentum transport by small scale gravity waves, which tend to be more strongly damped by finite volume numerical schemes at equivalent resolution, and are more accurately conserving angular momentum. In terms of the age distributions, finite volume numerical methods are by construction designed to provide a conservative representation of tracer transport. The spread between the age structure in the spectral and finite volume cores is likely due to differences in isentropic mixing across the barriers to transport in the subtropical stratosphere, and around the polar vortex. Future work will seek to more carefully attribute these differences to differences in resolved stirring (the filementation of tracer gradients by breaking waves) and numerical diffusion on the grid scale.

Appendix A : Specifying Tropical Winds

Specifying easterly in the tropical stratosphere is similar in spirit to specifying QBO winds in comprehensive climate models (e.g., Matthes et al. 2010). The key differences is that we specify a fixed easterly throughout out integrations, instead of an oscillating wind pattern. The tropical winds are nudged to a prescribed wind profile,

$$\partial_t u(\lambda, \phi, p, t) = -\frac{u(\lambda, \phi, p, t) - u_{eq}(p)}{\tau(\phi, p)}. \tag{5}$$

The wind profile is

$$u_{eq}(p) = \begin{cases} 0 & 0 \leq p \leq 0.3\ hPa \\ u_{1hPa} + \left(\frac{u_{sponge} - u_{1hPa}}{log(0.3)}\right) log(p) & 0.3\ hPa \leq p \leq 1\ hPa \\ u_{200hPa} + \left(\frac{u_{200hPa} - u_{1hPa}}{log(200)}\right) (log(p) - log(200)) + 10 sin\left(\frac{\pi log(p)}{log(200)}\right) & 1\ hPa \leq p \leq 200\ hPa \end{cases} \quad (6)$$

where $u_{sponge} = 0\ m/s$, $u_{1hPa} = -65\ m/s$ and $u_{200hPa} = -10\ m/s$. The resulting profile u_{eq} is plotted (in solid black) in Figure 2.

The damping extends from -15° to 15° latitude, and from 200 hPa to the model top. It is strongest at the equator, with a time scale $\tau = 40$ days, which is equal to the thermal relaxation timescale in the diabatic forcing. Above 3hPa, the damping is increased to a time scale of 10 days; this was found necessary, by trial and error, to keep control of the winds in the upper stratosphere. Away from the equator, the damping coefficient $1/\tau$ decreases as a Gaussian with a standard deviation of $\sigma = 5^\circ$. The analytical expression for the damping coefficient is

$$\tau = A(\phi, p)\tau_{max} \quad (7)$$

where

$$A(\phi, p) = \left[\frac{1}{2}\left(1 + tanh[-0.1(p - 200)] + 3\left(1 + tanh[-2\,(p-3)]\right)\right)\right] e^{\frac{-\phi^2}{2\sigma^2}}, \quad (8)$$

ϕ is the latitude, and p is the pressure level.

Appendix B : Model Description and Advection Schemes

The four cores employ very different numerics and discretizations to govern their dynamics (i.e., for the transport of heat and momentum), in addition to different numerical schemes for tracer advection. In addition, scale-separation between horizontal vs. vertical motion motivates the use of different schemes for horizontal and vertical transport. Table 3 provides a summary of the advection schemes employed by the dycores to transport the clock tracer.

Tracer Advection Schemes: The more recently developed dynamical cores (GFDL-FV3 and CAM-SE) use a vertically Lagrangian finite volume transport scheme by Lin and Rood (1996) in the vertical (identical to the one used for momentum advection), which is accompanied by regular remapping of the advected felds to fixed pressure levels after every certain timesteps. The remapping algorithm varies across dycores and are highlighted in the table. Among the relatively older cores, both CAM-PS and CAM-FV accomplish vertical advection using piecewise parabolic finite volume methods (PPM) (Colella and Woodward 1984).

For tracer advection in the horizontal, the FV3 core uses a positive-definite, grid based, finite-volume scheme proposed in Lin and Rood (1996). The advection is further constrained by numerical constraints proposed by Huynh (1997) for improved wave-resolution. Different from GFDL-FV3, the CAM-SE uses a strong stability preserving (SSP) fourth-order Runge-Kutta (RK4) method to time-march the transport equation. Among the older cores, the CAM-FV uses piecewise parabolic method (PPM) for horizontal tracer advection (identical to vertical transport) and the PS core uses a default spectral scheme to advect tracers in the spectral space.

Model resolution: We finally breifly explain the model resolution designations shown in Table 2. In the horizontal, we sought grids with comparable resolutions for the different dynamical cores. For the GFDL-PS core, we work with **T42** and **T85** truncation (**T**riangular truncation up to **42** and **85** spherical harmonics, respectively) that correspond to a 64×128 and 128×256 Gaussian grid respectively. For the GFDL-FV3 core, we work with **C48** and **C90** horizontal native grids (6 tiles with $\mathbf{48} \times \mathbf{48}$ and $\mathbf{90} \times \mathbf{90}$ points on the local grid) that are ultimately mapped onto a 96×192 and 180×360 latitude-longitude grid respectively using a 2nd order mass and energy conserving scheme. For the CAM-FV dycore, we use the F**19** and F**09** grids with 96×144 ($\mathbf{1.9^\circ} \times 2.5^\circ$) and 192×288 ($\mathbf{0.9^\circ} \times 1.25^\circ$) points on the lat-lon grid. Finally, for the CAM-SE dycore, we use the **NE16 NP4** and **NE30 NP4** grids: $\mathbf{16} \times \mathbf{16}$ and $\mathbf{30} \times \mathbf{30}$ native grid with $\mathbf{4} \times \mathbf{4}$ spectral element within each cell. In the nomenclature of the GFDL-FV3, this would be C64 and C120. These grids are interpolated on to a considerably finer 129×256 and 257×512 latitude-longitude grid, respectively, using bilinear interpolation.

In the vertical, we use 40 levels for the coarse runs and refine it to 80 levels for high resolution runs. To select the vertical levels, we use angular momentum conserving σ-pressure hybrid coordinate (Simmons and Burridge 1981) - which is a combination of pure σ (terrain following) coordinates following the orography upto 500hPa and pure pressure above 200hPa. The intermediate region between 200hPa and 500hPa ensures smooth linear transition from the terrain following coordinates below to pure pressure coordinates above. The mean pressure levels in the stratosphere are decided using the equations provided in paragraph 21 of PK02.

Acknowledgement

We acknowledge support of the US National Science Foundation through grant AGS-1546585 to New York University. We also thank Marianna Linz, Alan Plumb, and Olivier Pauluis for insightful discussion and constructive criticism of the research.

References

D. G. Andrews and M. E. McIntyre. Generalized Eliassen-Palm and Charney-Drazin Theorems for Waves on Axismmetric Mean Flows in Compressible Atmospheres. *J. Atmos. Sci.*, 35(2):175–185, February 1978. ISSN 0022-4928. doi: 10.1175/1520-0469(1978)035⟨0175:GEPACD⟩2.0.CO;2.

D. G. Andrews and M. F. McIntyre. Planetary Waves in Horizontal and Vertical Shear: Asymptotic Theory for Equatorial Waves in Weak Shear. *J. Atmos. Sci.*, 33(11):2049–2053, November 1976. ISSN 0022-4928. doi: 10.1175/1520-0469(1976)033⟨2049:PWIHAV⟩2.0.CO;2.

Elizabeth A. Barnes and Lorenzo Polvani. Response of the Midlatitude Jets, and of Their Variability, to Increased Greenhouse Gases in the CMIP5 Models. *J. Climate*, 26(18):7117–7135, April 2013. ISSN 0894-8755. doi: 10.1175/JCLI-D-12-00536.1.

John P. Boyd. The Noninteraction of Waves with the Zonally Averaged Flow on a Spherical Earth and the Interrelationships on Eddy Fluxes of Energy, Heat and Momentum. *J. Atmos. Sci.*, 33(12):2285–2291, December 1976. ISSN 0022-4928. doi: 10.1175/1520-0469(1976)033⟨2285:TNOWWT⟩2.0.CO;2.

A. W. Brewer. Evidence for a world circulation provided by the measurements of helium and water vapour distribution in the stratosphere. *Quarterly Journal of the Royal Meteorological Society*, 75(326):351–363, 1949. ISSN 1477-870X. doi: 10.1002/qj.49707532603.

Neal Butchart. The Brewer-Dobson circulation. *Reviews of Geophysics*, 52(2):157–184, 2014. ISSN 1944-9208. doi: 10.1002/2013RG000448.

Amy H. Butler, David W. J. Thompson, and Ross Heikes. The Steady-State Atmospheric Circulation Response to Climate Change–like Thermal Forcings in a Simple General Circulation Model. *J. Climate*, 23(13):3474–3496, February 2010. ISSN 0894-8755. doi: 10.1175/2010JCLI3228.1.

Amy H. Butler, Jeremiah P. Sjoberg, Dian J. Seidel, and Karen H. Rosenlof. A sudden stratospheric warming compendium. *Earth System Science Data*, 9(1): 63–76, February 2017. ISSN 1866-3508. doi: https://doi.org/10.5194/essd-9-63-2017.

Richard L. Carpenter, Kelvin K. Droegemeier, Paul R. Woodward, and Carl E. Hane. Application of the Piecewise Parabolic Method (PPM) to Meteorological Modeling. *Mon. Wea. Rev.*, 118(3):586–612, March 1990. ISSN 0027-0644. doi: 10.1175/1520-0493(1990)118⟨0586:AOTPPM⟩2.0.CO;2.

Andrew J. Charlton-Perez, Mark P. Baldwin, Thomas Birner, Robert X. Black, Amy H. Butler, Natalia Calvo, Nicholas A. Davis, Edwin P. Gerber, Nathan Gillett, Steven Hardiman, Junsu Kim, Kirstin Krüger, Yun-Young Lee, Elisa Manzini, Brent A. McDaniel, Lorenzo Polvani, Thomas Reichler, Tiffany A. Shaw, Michael Sigmond, Seok-Woo Son, Matthew Toohey, Laura Wilcox, Shigeo Yoden, Bo Christiansen, François Lott, Drew Shindell, Seiji Yukimoto, and Shingo Watanabe. On the lack of stratospheric dynamical variability in low-top versions of the CMIP5 models. *Journal of Geophysical Research: Atmospheres*, 118(6):2494–2505, 2013. ISSN 2169-8996. doi: 10.1002/jgrd.50125.

Phillip Colella and Paul R Woodward. The Piecewise Parabolic Method (PPM) for gas-dynamical simulations. *Journal of Computational Physics*, 54(1): 174–201, April 1984. ISSN 0021-9991. doi: 10.1016/0021-9991(84)90143-8.

William J. Collins, Jean-François Lamarque, Michael Schulz, Olivier Boucher, Veronika Eyring, Michaela I. Hegglin, Amanda Maycock, Gunnar Myhre, Michael Prather, Drew Shindell, and Steven J. Smith. AerChemMIP: Quantifying the effects of chemistry and aerosols in CMIP6. *Geoscientific Model Development*, 10(2):585–607, February 2017. ISSN 1991-959X. doi: https://doi.org/10.5194/gmd-10-585-2017.

John M. Dennis, Jim Edwards, Katherine J. Evans, Oksana Guba, Peter H. Lauritzen, Arthur A. Mirin, Amik St-Cyr, Mark A. Taylor, and Patrick H. Worley. CAM-SE: A scalable spectral element dynamical core for the Community Atmosphere Model. *The International Journal of High Performance Computing Applications*, 26(1):74–89, February 2012. ISSN 1094-3420. doi: 10.1177/1094342011428142.

GMB Dobson. Origin and distribution of the polyatomic molecules in the atmosphere. *Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences*, 236(1205):187–193, August 1956. doi: 10.1098/rspa.1956.0127.

GMB Dobson, Harrison D. N., and Lindemann Frederick Alexander. Measurements of the amount of ozone in the earth's atmosphere and its relation to other geophysical conditions. *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character*, 110(756): 660–693, April 1926. doi: 10.1098/rspa.1926.0040.

A. Engel, T. Möbius, H. Bönisch, U. Schmidt, R. Heinz, I. Levin, E. Atlas, S. Aoki, T. Nakazawa, S. Sugawara, F. Moore, D. Hurst, J. Elkins, S. Schauffler, A. Andrews, and K. Boering. Age of stratospheric air unchanged within uncertainties over the past 30 years. *Nature Geoscience*, 2(1):28–31, January 2009. ISSN 1752-0908. doi: 10.1038/ngeo388.

V. Eyring, S. Bony, G. A. Meehl, C. A. Senior, B. Stevens, R. J. Stouffer, and K. E. Taylor. Overview of the coupled model intercomparison project phase 6 (cmip6) experimental design and organization. *Geoscientific Model Development*, 9(5):1937–1958, 2016. doi: 10.5194/gmd-9-1937-2016.

Chaim I. Garfinkel, Tiffany A. Shaw, Dennis L. Hartmann, and Darryn W. Waugh. Does the Holton–Tan Mechanism Explain How the Quasi-Biennial Oscillation Modulates the Arctic Polar Vortex? *J. Atmos. Sci.*, 69(5):1713–1733, February 2012. ISSN 0022-4928. doi: 10.1175/JAS-D-11-0209.1.

Edwin P. Gerber and Lorenzo M. Polvani. Stratosphere–Troposphere Coupling in a Relatively Simple AGCM: The Importance of Stratospheric Variability. *J. Climate*, 22(8):1920–1933, April 2009. ISSN 0894-8755. doi: 10.1175/2008JCLI2548.1.

Edwin P. Gerber, Amy Butler, Natalia Calvo, Andrew Charlton-Perez, Marco Giorgetta, Elisa Manzini, Judith Perlwitz, Lorenzo M. Polvani, Fabrizio Sassi, Adam A. Scaife, Tiffany A. Shaw, Seok-Woo Son, and Shingo Watanabe. Assessing and Understanding the Impact of Stratospheric Dynamics and Variability on the Earth System. *Bull. Amer. Meteor. Soc.*, 93(6):845–859, January 2012. ISSN 0003-0007. doi: 10.1175/BAMS-D-11-00145.1.

Marco A. Giorgetta, Elisa Manzini, and Erich Roeckner. Forcing of the quasi-biennial oscillation from a broad spectrum of atmospheric waves. *Geophysical Research Letters*, 29(8):86–1–86–4, 2002. ISSN 1944-8007. doi: 10.1029/2002GL014756.

Timothy M. Hall and R. Alan Plumb. Age as a diagnostic of stratospheric transport. *J. Geophys. Res.*, 99(D1):1059, 1994. ISSN 0148-0227. doi: 10.1029/93JD03192.

Timothy M. Hall, Darryn W. Waugh, Kristie A. Boering, and R. Alan Plumb. Evaluation of transport in stratospheric models. *Journal of Geophysical Research: Atmospheres*, 104(D15):18815–18839, 1999. ISSN 2156-2202. doi: 10.1029/1999JD900226.

Jian Hang, Mats Sandberg, and Yuguo Li. Age of air and air exchange efficiency in idealized city models. *Building and Environment*, 44(8):1714–1723, August 2009. ISSN 0360-1323. doi: 10.1016/j.buildenv.2008.11.013.

P. H. Haynes, M. E. McIntyre, T. G. Shepherd, C. J. Marks, and K. P. Shine. On the "Downward Control" of Extratropical Diabatic Circulations by Eddy-Induced Mean Zonal Forces. *J. Atmos. Sci.*, 48(4):651–678, February 1991. ISSN 0022-4928. doi: 10.1175/1520-0469(1991)048⟨0651:OTCOED⟩2.0.CO;2.

Isaac M. Held and Max J. Suarez. A Proposal for the Intercomparison of the Dynamical Cores of Atmospheric General Circulation Models. *Bull. Amer. Meteor. Soc.*, 75(10):1825–1830, October 1994. ISSN 0003-0007. doi: 10.1175/1520-0477(1994)075⟨1825:APFTIO⟩2.0.CO;2.

James R. Holton and Hsiu-Chi Tan. The Influence of the Equatorial Quasi-Biennial Oscillation on the Global Circulation at 50 mb. *J. Atmos. Sci.*, 37(10): 2200–2208, October 1980. ISSN 0022-4928. doi: 10.1175/1520-0469(1980)037⟨2200:TIOTEQ⟩2.0.CO;2.

James R. Holton, Peter H. Haynes, Michael E. McIntyre, Anne R. Douglass, Richard B. Rood, and Leonhard Pfister. Stratosphere-troposphere exchange. *Reviews of Geophysics*, 33(4):403–439, 1995. ISSN 1944-9208. doi: 10.1029/95RG02097.

H. T. Huynh. Schemes and constraints for advection. In Paul Kutler, Jolen Flores, and Jean-Jacques Chattot, editors, *Fifteenth International Conference on Numerical Methods in Fluid Dynamics*, volume 490, pages 498–503. Springer Berlin Heidelberg, 1997. ISBN 978-3-540-63054-8. doi: 10.1007/BFb0107151.

Alexey Yu. Karpechko, Douglas Maraun, and Veronika Eyring. Improving Antarctic Total Ozone Projections by a Process-Oriented Multiple Diagnostic Ensemble Regression. *J. Atmos. Sci.*, 70(12):3959–3976, August 2013. ISSN 0022-4928. doi: 10.1175/JAS-D-13-071.1.

James Kent, Paul A. Ullrich, and Christiane Jablonowski. Dynamical core model intercomparison project: Tracer transport test cases. *Quarterly Journal of the Royal Meteorological Society*, 140(681):1279–1293, 2014. ISSN 1477-870X. doi: 10.1002/qj.2208.

Hideji Kida. General Circulation of Air Parcels and Transport Characteristics Derived from a Hemispheric GCM Part 2. Very Long-Term Motions of Air Parcels in the Troposphere and Stratosphere. *Journal of the Meteorological Society of Japan*, 61(4):14, 1983.

Joseph Kidston, Adam A. Scaife, Steven C. Hardiman, Daniel M. Mitchell, Neal Butchart, Mark P. Baldwin, and Lesley J. Gray. Stratospheric influence on tropospheric jet streams, storm tracks and surface weather. *Nature Geosci*, 8(6):433–440, June 2015. ISSN 1752-0894, 1752-0908. doi: 10.1038/ngeo2424.

Ingemar P. E. Kinnmark and William G. Gray. One step integration methods with maximum stability regions. *Mathematics and Computers in Simulation*, 26 (2):87–92, April 1984. ISSN 0378-4754. doi: 10.1016/0378-4754(84)90039-9.

P. H. Lauritzen, R. D. Nair, A. R. Herrington, P. Callaghan, S. Goldhaber, J. M. Dennis, J. T. Bacmeister, B. E. Eaton, C. M. Zarzycki, Mark A. Taylor, P. A. Ullrich, T. Dubos, A. Gettelman, R. B. Neale, B. Dobbins, K. A. Reed, C. Hannay, B. Medeiros, J. J. Benedict, and J. J. Tribbia. NCAR Release of CAM-SE in CESM2.0: A Reformulation of the Spectral Element Dynamical Core in Dry-Mass Vertical Coordinates With Comprehensive Treatment of Condensates and Energy. *Journal of Advances in Modeling Earth Systems*, 10(7):1537–1570, 2018. ISSN 1942-2466. doi: 10.1029/2017MS001257.

Peter H. Lauritzen, Arthur A. Mirin, John Truesdale, Kevin Raeder, Jeffrey L. Anderson, Julio Bacmeister, and Richard B. Neale. Implementation of new diffusion/filtering operators in the CAM-FV dynamical core. *The International Journal of High Performance Computing Applications*, 26(1):63–73, February 2012. ISSN 1094-3420. doi: 10.1177/1094342011410088.

Peter H. Lauritzen, Julio T. Bacmeister, Thomas Dubos, Sébastien Lebonnois, and Mark A. Taylor. Held-Suarez simulations with the Community Atmosphere Model Spectral Element (CAM-SE) dynamical core: A global axial angular momentum analysis using Eulerian and floating Lagrangian vertical coordinates. *Journal of Advances in Modeling Earth Systems*, 6(1):129–140, 2014. ISSN 1942-2466. doi: 10.1002/2013MS000268.

P.H. Lauritzen, C. Jablonowski, M.A. Taylor, and R. D. Nair. *Numerical Techniques for Global Atmospheric Models*, volume 80 of *Lecture Notes in Computational Science and Engineering*. Springer-Verlag Berlin Heidelberg, 1 edition, 2011. ISBN 978-3-642-11639-1.

Shian-Jiann Lin. A “Vertically Lagrangian” Finite-Volume Dynamical Core for Global Models. *Mon. Wea. Rev.*, 132(10):2293–2307, October 2004. ISSN 0027-0644. doi: 10.1175/1520-0493(2004)132⟨2293:AVLFDC⟩2.0.CO;2.

Shian-Jiann Lin and Richard B. Rood. Multidimensional Flux-Form Semi-Lagrangian Transport Schemes. *Mon. Wea. Rev.*, 124(9):2046–2070, September 1996. ISSN 0027-0644. doi: 10.1175/1520-0493(1996)124⟨2046:MFFSLT⟩2.0.CO;2.

Marianna Linz, R. Alan Plumb, Edwin P. Gerber, and Aditi Sheshadri. The Relationship between Age of Air and the Diabatic Circulation of the Stratosphere. *J. Atmos. Sci.*, 73(11):4507–4518, August 2016. ISSN 0022-4928. doi: 10.1175/JAS-D-16-0125.1.

Marianna Linz, R. Alan Plumb, Edwin P. Gerber, Florian J. Haenel, Gabriele Stiller, Douglas E. Kinnison, Alison Ming, and Jessica L. Neu. The strength of the meridional overturning circulation of the stratosphere. *Nature Geosci*, 10(9):663–667, September 2017. ISSN 1752-0894, 1752-0908. doi: 10.1038/ngeo3013.

David J. Lorenz and Eric T. DeWeaver. Tropopause height and zonal wind response to global warming in the IPCC scenario integrations. *Journal of Geophysical Research: Atmospheres*, 112(D10), 2007. ISSN 2156-2202. doi: 10.1029/2006JD008087.

Katja Matthes, Daniel R. Marsh, Rolando R. Garcia, Douglas E. Kinnison, Fabrizio Sassi, and Stacy Walters. Role of the QBO in modulating the influence of the 11 year solar cycle on the atmosphere using constant forcings. *Journal of Geophysical Research: Atmospheres*, 115(D18), 2010. ISSN 2156-2202. doi: 10.1029/2009JD013020.

R. M. Mcinturff. Stratospheric warmings: Synoptic, dynamic and general-circulation aspects. Technical report, January 1978.

M. E. McIntyre and T. N. Palmer. Breaking planetary waves in the stratosphere. *Nature*, 305(5935):593–600, October 1983. ISSN 0028-0836, 1476-4687. doi: 10.1038/305593a0.

M. E McIntyre and T. N Palmer. The 'surf zone' in the stratosphere. *Journal of Atmospheric and Terrestrial Physics*, 46(9):825–849, September 1984. ISSN 0021-9169. doi: 10.1016/0021-9169(84)90063-1.

B. M. Monge-Sanz, M. P. Chipperfield, A. J. Simmons, and S. M. Uppala. Mean age of air and transport in a CTM: Comparison of different ECMWF analyses. *Geophysical Research Letters*, 34(4), 2007. ISSN 1944-8007. doi: 10.1029/2006GL028515.

R. J. Murgatroyd and F. Singleton. Possible meridional circulations in the stratosphere and mesosphere. *Quarterly Journal of the Royal Meteorological Society*, 87(372):125–135, 1961. ISSN 1477-870X. doi: 10.1002/qj.49708737202.

Jessica L. Neu and R. Alan Plumb. Age of air in a "leaky pipe" model of stratospheric transport. *Journal of Geophysical Research: Atmospheres*, 104(D16): 19243–19255, 1999. ISSN 2156-2202. doi: 10.1029/1999JD900251.

Clara Orbe, Luke D. Oman, Susan E. Strahan, Darryn W. Waugh, Steven Pawson, Lawrence L. Takacs, and Andrea M. Molod. Large-Scale Atmospheric Transport in GEOS Replay Simulations. *Journal of Advances in Modeling Earth Systems*, 9(7):2545–2560, 2017a. ISSN 1942-2466. doi: 10.1002/2017MS001053.

Clara Orbe, Darryn W. Waugh, Huang Yang, Jean-Francois Lamarque, Simone Tilmes, and Douglas E. Kinnison. Tropospheric transport differences between models using the same large-scale meteorological fields. *Geophysical Research Letters*, 44(2):1068–1078, 2017b. ISSN 1944-8007. doi: 10.1002/2016GL071339.

Olivier Pauluis, Arnaud Czaja, and Robert Korty. The Global Atmospheric Circulation in Moist Isentropic Coordinates. *J. Climate*, 23(11):3077–3093, December 2009. ISSN 0894-8755. doi: 10.1175/2009JCLI2789.1.

R. Alan Plumb. Stratospheric Transport. *Journal of the Meteorological Society of Japan. Ser. II*, 80(4B):793–809, 2002. doi: 10.2151/jmsj.80.793.

R. Alan Plumb and Robert C. Bell. A model of the quasi-biennial oscillation on an equatorial beta-plane. *Quarterly Journal of the Royal Meteorological Society*, 108(456):335–352, 1982. ISSN 1477-870X. doi: 10.1002/qj.49710845604.

Lorenzo M. Polvani and Paul J. Kushner. Tropospheric response to stratospheric perturbations in a relatively simple general circulation model. *Geophysical Research Letters*, 29(7):18–1–18–4, 2002. ISSN 1944-8007. doi: 10.1029/2001GL014284.

William M. Putman and Shian-Jiann Lin. Finite-volume transport on various cubed-sphere grids. *Journal of Computational Physics*, 227(1):55–78, November 2007. ISSN 0021-9991. doi: 10.1016/j.jcp.2007.07.022.

Richard B. Rood. Numerical advection algorithms and their role in atmospheric transport and chemistry models. *Reviews of Geophysics*, 25(1):71–100, 1987. ISSN 1944-9208. doi: 10.1029/RG025i001p00071.

Emily Shuckburgh, Warwick Norton, Alan Iwi, and Peter Haynes. Influence of the quasi-biennial oscillation on isentropic transport and mixing in the tropics and subtropics. *Journal of Geophysical Research: Atmospheres*, 106(D13):14327–14337, 2001. ISSN 2156-2202. doi: 10.1029/2000JD900664.

A. J. Simmons and D. M. Burridge. An Energy and Angular-Momentum Conserving Vertical Finite-Difference Scheme and Hybrid Vertical Coordinates. *Mon. Wea. Rev.*, 109(4):758–766, April 1981. ISSN 0027-0644. doi: 10.1175/1520-0493(1981)109⟨0758:AEAAMC⟩2.0.CO;2.

A. H. Sobel, R. A. Plumb, and D. W. Waugh. Methods of Calculating Transport across the Polar Vortex Edge. *J. Atmos. Sci.*, 54(18):2241–2260, September 1997. ISSN 0022-4928. doi: 10.1175/1520-0469(1997)054⟨2241:MOCTAT⟩2.0.CO;2.

Seok-Woo Son and Sukyoung Lee. The Response of Westerly Jets to Thermal Driving in a Primitive Equation Model. *J. Atmos. Sci.*, 62(10):3741–3757, October 2005. ISSN 0022-4928. doi: 10.1175/JAS3571.1.

Ronald D. Townsend and Donald R. Johnson. A Diagnostic Study of the Isentropic Zonally Averaged Mass Circulation during the First GARP Global Experiment. *J. Atmos. Sci.*, 42(15):1565–1579, August 1985. ISSN 0022-4928. doi: 10.1175/1520-0469(1985)042⟨1565:ADSOTI⟩2.0.CO;2.

G. K. Vallis, G. Colyer, R. Geen, E. P. Gerber, M. Jucker, P. Maher, A. Paterson, M. Pietschnig, J. Penn, and S. I. Thomson. Isca, v1.0: A framework for the global modelling of the atmospheres of earth and other planets at varying levels of complexity. 11:843–859, 2018. doi: 10.5194/gmd-11-843-2018.

Darryn Waugh and Timothy Hall. Age of Stratospheric Air: Theory, Observations, and Models. *Reviews of Geophysics*, 40(4):1–1–1–26, 2002. ISSN 1944-9208. doi: 10.1029/2000RG000101.

Sabrina Wenzel, Veronika Eyring, Edwin P. Gerber, and Alexey Yu. Karpechko. Constraining Future Summer Austral Jet Stream Positions in the CMIP5 Ensemble by Process-Oriented Multiple Diagnostic Regression. *J. Climate*, 29(2):673–687, January 2016. ISSN 0894-8755. doi: 10.1175/JCLI-D-15-0412.1.

David L. Williamson. The Evolution of Dynamical Cores for Global Atmospheric Models. *Journal of the Meteorological Society of Japan. Ser. II*, 85B:241–269, 2007. doi: 10.2151/jmsj.85B.241.

David L. Williamson, John B. Drake, James J. Hack, Rüdiger Jakob, and Paul N. Swarztrauber. A standard test set for numerical approximations to the shallow water equations in spherical geometry. *Journal of Computational Physics*, 102(1):211–224, September 1992. ISSN 0021-9991. doi: 10.1016/S0021-9991(05)80016-6.

Xiaokang Wu, Huang Yang, Darryn W. Waugh, Clara Orbe, Simone Tilmes, and Jean-Francois Lamarque. Spatial and temporal variability of interhemispheric transport times. *Atmospheric Chemistry and Physics*, 18(10):7439–7452, May 2018. ISSN 1680-7316. doi: https://doi.org/10.5194/acp-18-7439-2018.

Weiye Yao and Christiane Jablonowski. Idealized Quasi-Biennial Oscillations in an Ensemble of Dry GCM Dynamical Cores. *J. Atmos. Sci.*, 72(6):2201–2226, June 2015. ISSN 0022-4928, 1520-0469. doi: 10.1175/JAS-D-14-0236.1.

Weiye Yao and Christiane Jablonowski. The Impact of GCM Dynamical Cores on Idealized Sudden Stratospheric Warmings and Their QBO Interactions. *J. Atmos. Sci.*, 73(9):3397–3421, June 2016. ISSN 0022-4928. doi: 10.1175/JAS-D-15-0242.1.

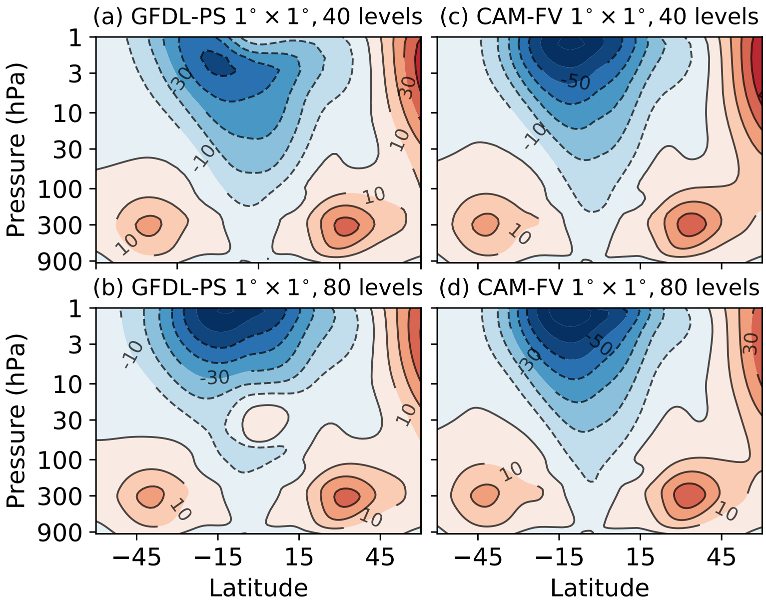

Figure 1. Zonal Winds $\bar{u}$ (in m/s) in 2 different dynamical cores : (left) pseudospectral (GFDL-PS) and (right) finite volume (CAM-FV) with 40 vertical levels (top) and 80 vertical levels (bottom). At high vertical resolution, pseudospectral core develops westerlies in the tropical stratosphere (20-80hPa). The finite volume core consistently resolve easterlies at both vertical resolutions. Both the cores have comparable horizontal resolution. Contour intervals : 10 m/s.

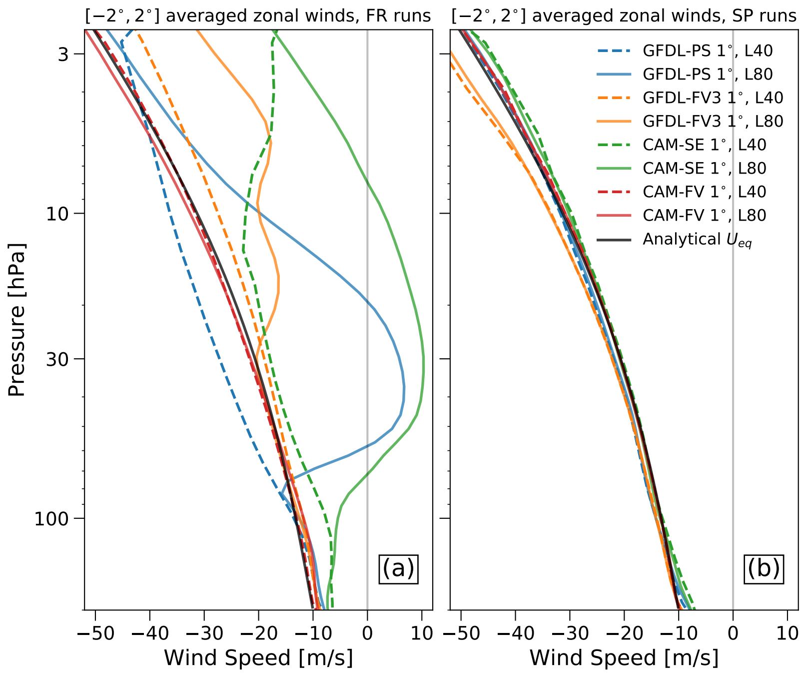

Figure 2. Zonal winds $\bar{u}$ averaged over $[-2°, 2°]$ for the 4 dynamical cores : GFDL-PS (blue), GFDL-FV3 (orange), CAM-SE (green) and CAM-FV (red), as obtained from the (left) FR integrations and (right) SP integrations. The solid curves and dashed curves show resolved winds at low (40) and high (80) vertical resolution respectively. The black line shows the analytical wind profile specified in the cores in the tropics (see Section 3).

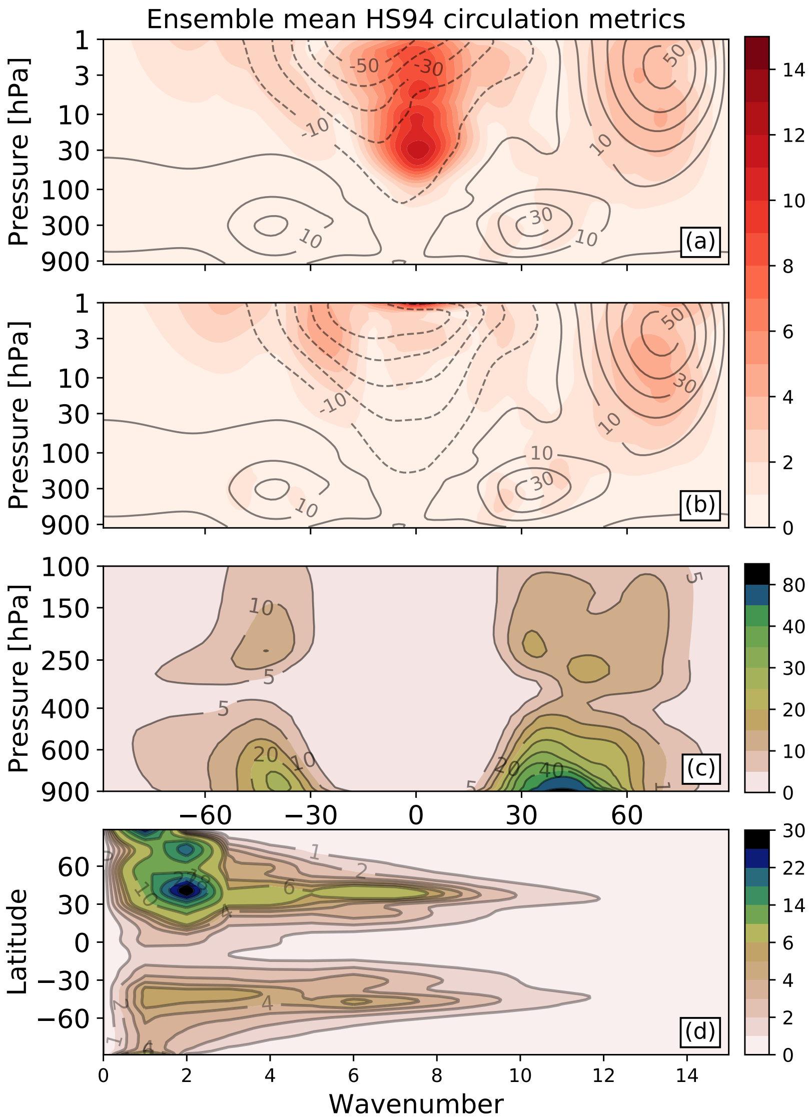

Figure 3. Ensemble mean Zonal mean zonal wind for the (a) free running and (b) specified tropical wind integrations. The black contours show the ensemble mean computed for the 1° L80 runs (T85L80, C90L80, NE30L80, F09L80) and the colors show the deviation in the winds from the ensemble mean. (c) The ensemble mean eddy temperature variance in the troposphere. (d) The ensemble mean vertically integrated zonal spectra for the eddy kinetic energy.

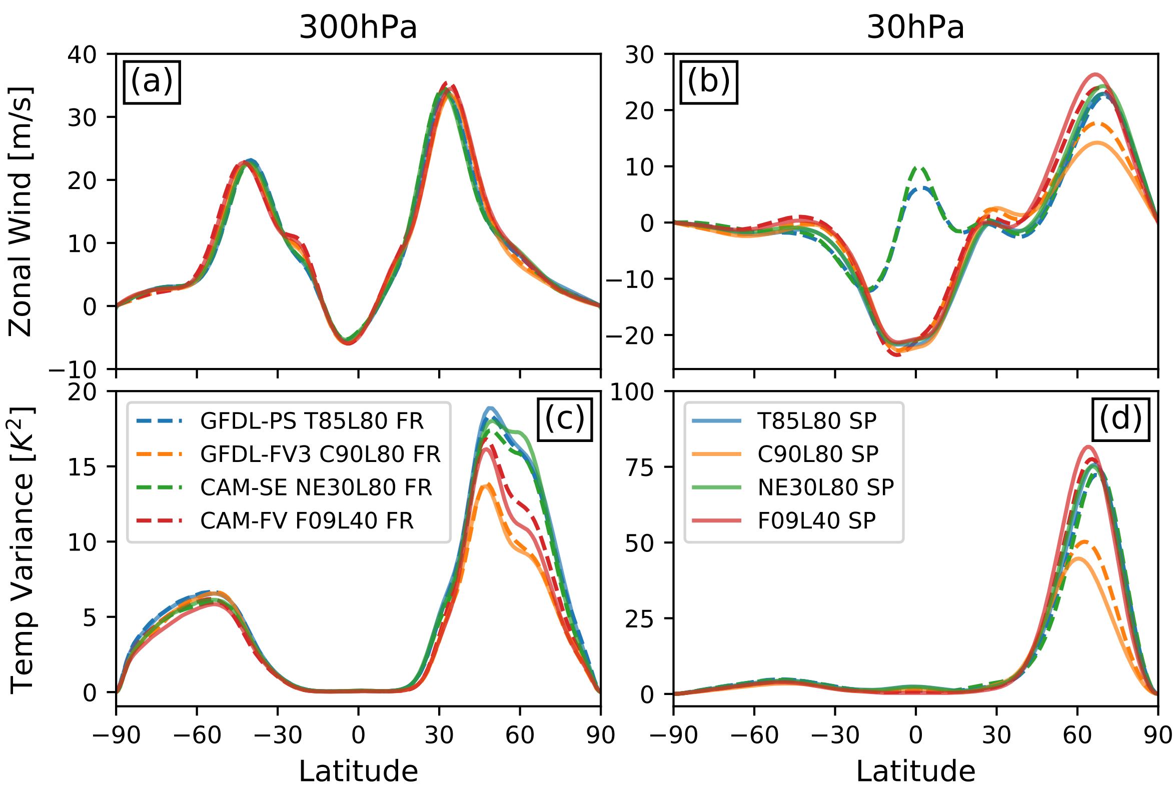

Figure 4. Zonal mean zonal wind (top; (a) and (b)) and eddy temperature variance (bottom; (c) and (d)) as in Figure 3(a) and (b) but at 2 different vertical levels for 1° L80 integrations from the four dycores. The left column compares the 2 quantities at 300hPa in the troposphere and the right column compares them at 30hPa in the stratosphere. The dashed and solid curves show the quantities for the free running (FR) and the specified tropical winds (SP) integrations respectively. The GFDL-pseudospectral core is shown in blue, GFDL-FV3 core in orange, CAM-spectral element in green and the CAM-finite volume in red.

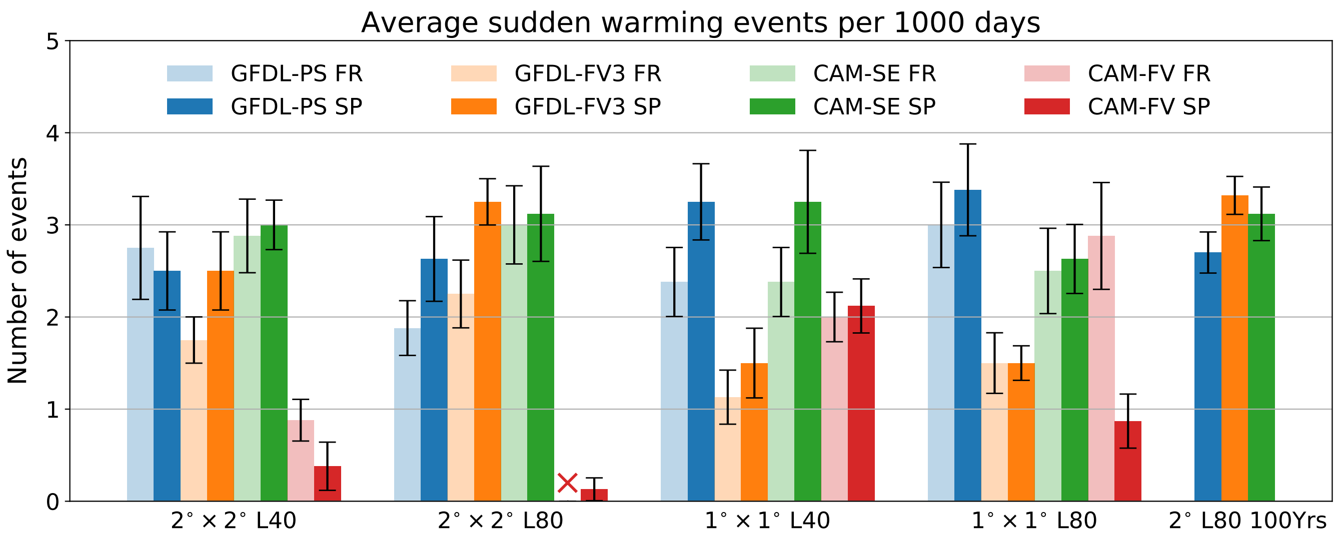

Figure 5. Average sudden warming frequencies for the dynamical cores grouped by horizontal and vertical resolutions during a 1000-day interval. The histograms represent averages taken over 8 independent samples and the error bars indicate the ensemble standard deviation for the mean frequency. For each resolution group, the lighter colored bars represent the free running (FR) integrations and the brighter colored bars represent the specified (SP) runs. The red cross in the second group indicates the absence of any sudden warming events for the F19L80 FR integration. For statistical significance, the 3 of the 4 cores were integrated at 2° L80 resolution 100 years. The statistics for these long integrations are shows by the rightmost group.

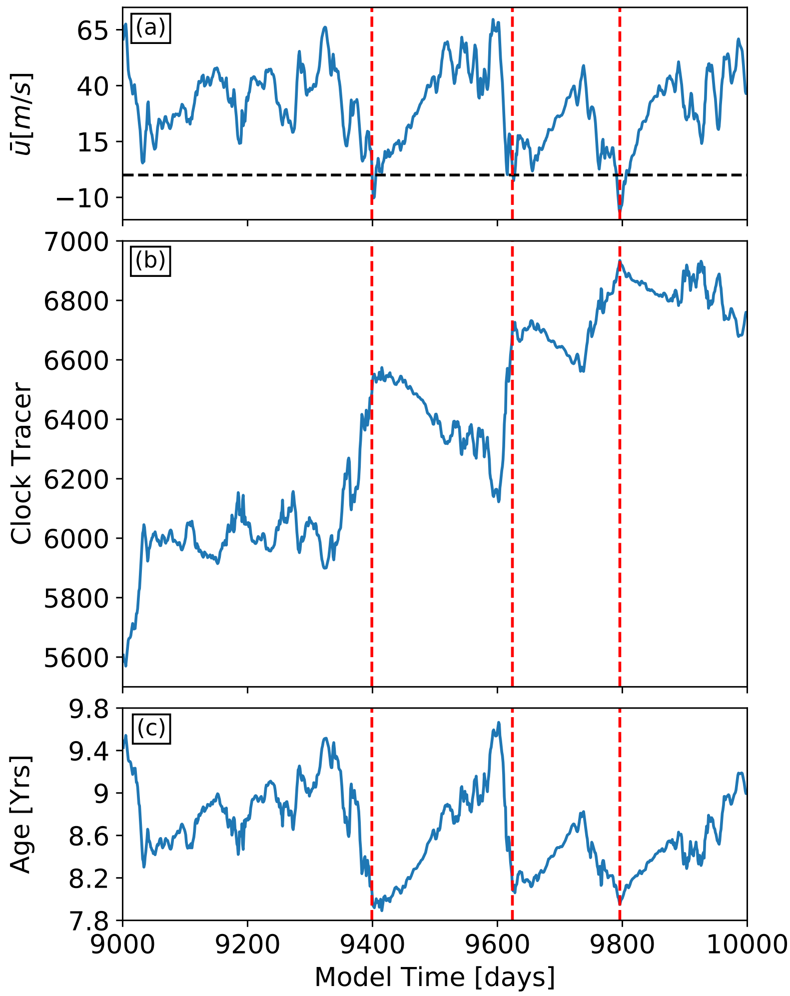

Figure 6. (a) Zonal mean zonal wind (in m/s) (b) clock tracer concentration and (c) age-of-air (in years) at 60N and 10hPa for a 1000-day interval for the CAM-SE NE30L80 run. Zonal wind reversal, from westerly to easterly, indicates major stratospheric sudden warming events characterized by temporary disintegration of the boreal winter polar vortex. These events are marked by a dashed red line around the 9400th, 9625th and 9800th model day.

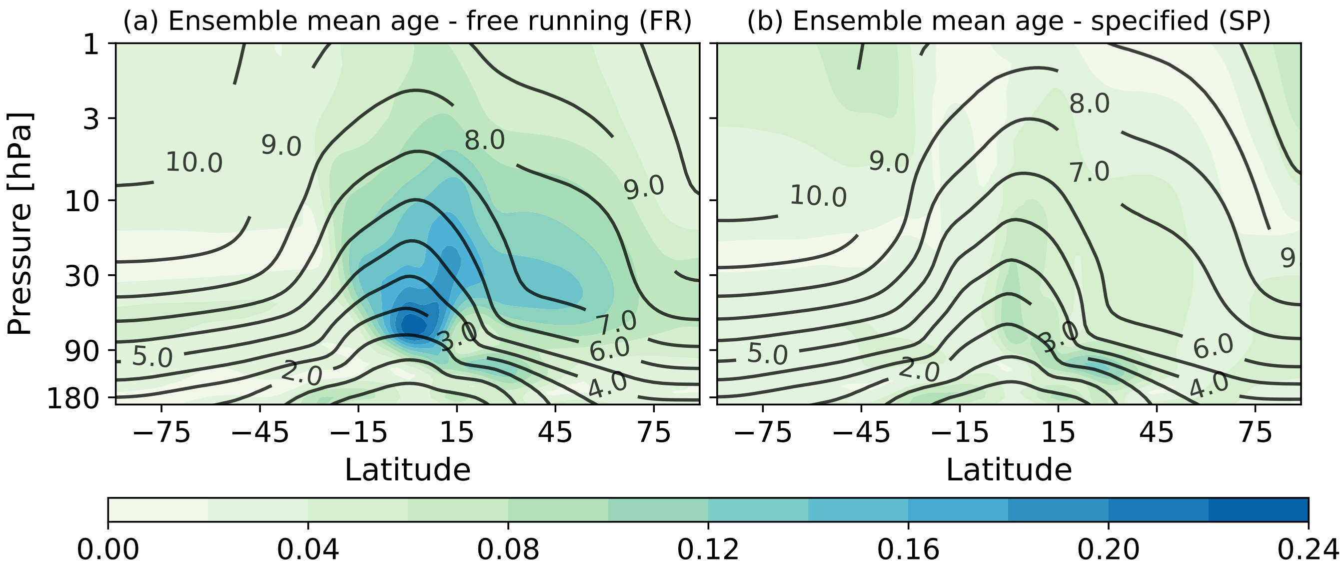

Figure 7. Ensemble mean age of air (years) in the stratosphere in solid contours for (a) the FR integrations and (b) the specified runs . The colors in the background show the fractional age deviation from the ensemble mean. A deviation of 0.1 at a position where the ensemble mean age is 10 yrs means a total standard deviation of 1 yr across the ensemble.

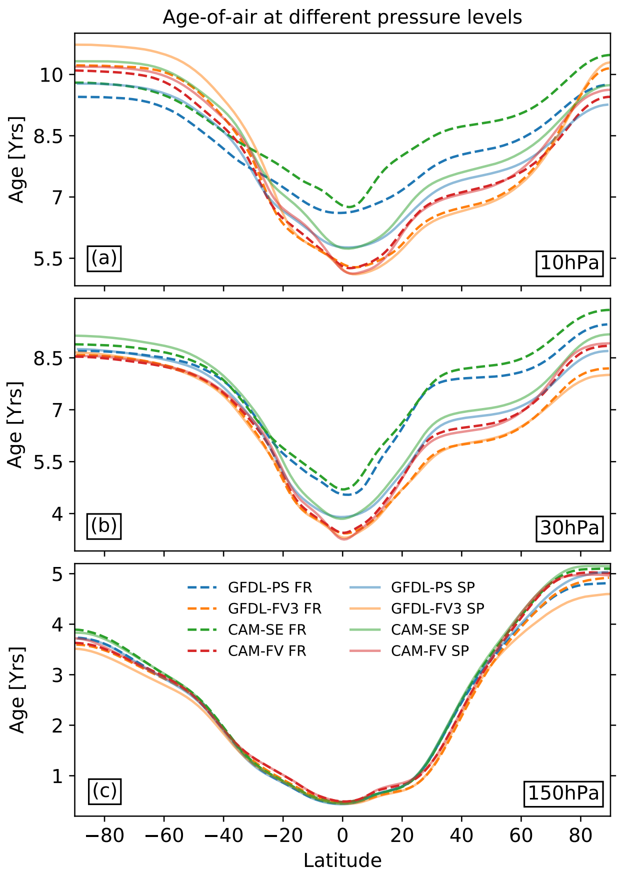

Figure 8. Age of air (in years) for the four dycores at 3 different pressure levels : (a) 10hPa (b) 30hPa and (c) 150hPa. FR and SP integrations are shown using dashed and solid curves respectively. The 3 levels are picked to represent the upper, middle and lower stratosphere respectively.

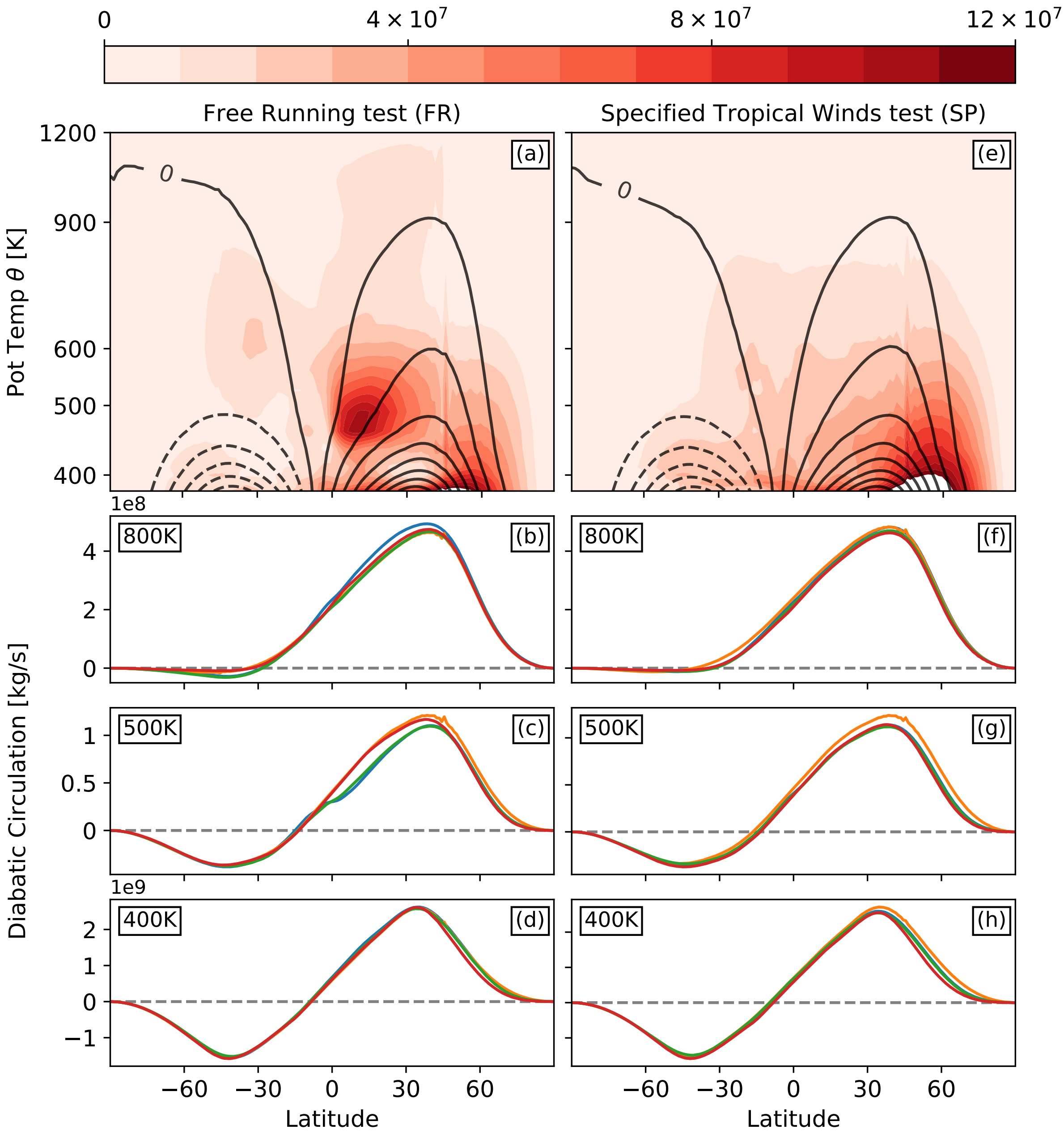

Figure 9. Ensemble averaged diabatic circulation in the stratosphere (in solid black) and ensemble deviation in circulation (in color) in potential temperature coordinates for (a) FR integrations and (e) SP integrations. (b)-(d) shows the diabatic circulation strength from the FR integrations of at 3 different isentropes : (b) 800K, (c) 500K and (d) 400K. (f)-(h) shows the same but for the SP integrations. As before, the most resolved runs were considered for the ensemble averaging with an exception of CAM-FV. The GFDL-pseudospectral is shown in blue, the GFDL-FV3 in orange, the CAM-SE in green and the CAM-FV in red. Contour interval for diabatic circulation is 0.4×10^9 kg/s.

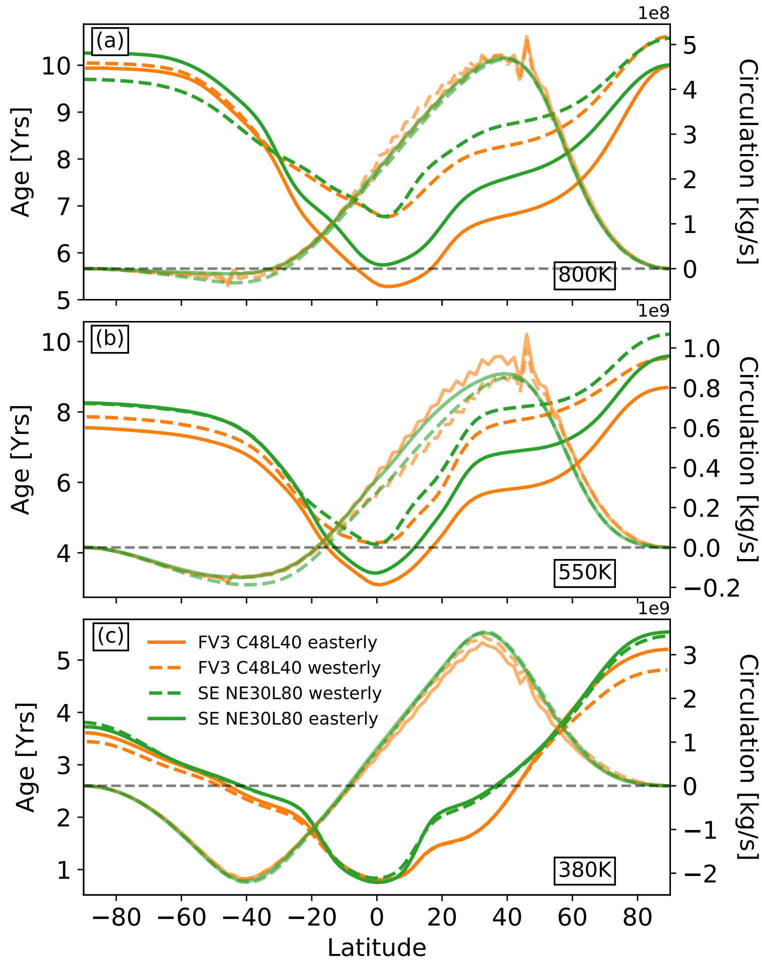

Figure 10. Age of air (dark curves; left axis) and diabatic circulation (lighter curves; right axis) at 3 different isentropic levels in the stratosphere : (a) 800K, (b) 550K and (c) 380K for the FR and SP CAM-SE and the FR FV3 and the westerly imposed FV3. The solid curves show the age and circulation for the CAM-SE and FV3 core with easterly winds i.e. NE30L80 (SP) and GFDL-FV3 (FR). The dashed curves show the same but for the runs with westerly winds i.e. NE30L80 (FR) and GFDL-FV3 (Specified westerly). The GFDL-FV3 is shown in orange and the CAM-SE in green. The dashed black line representing 0 diabatic circulation is added for reference.

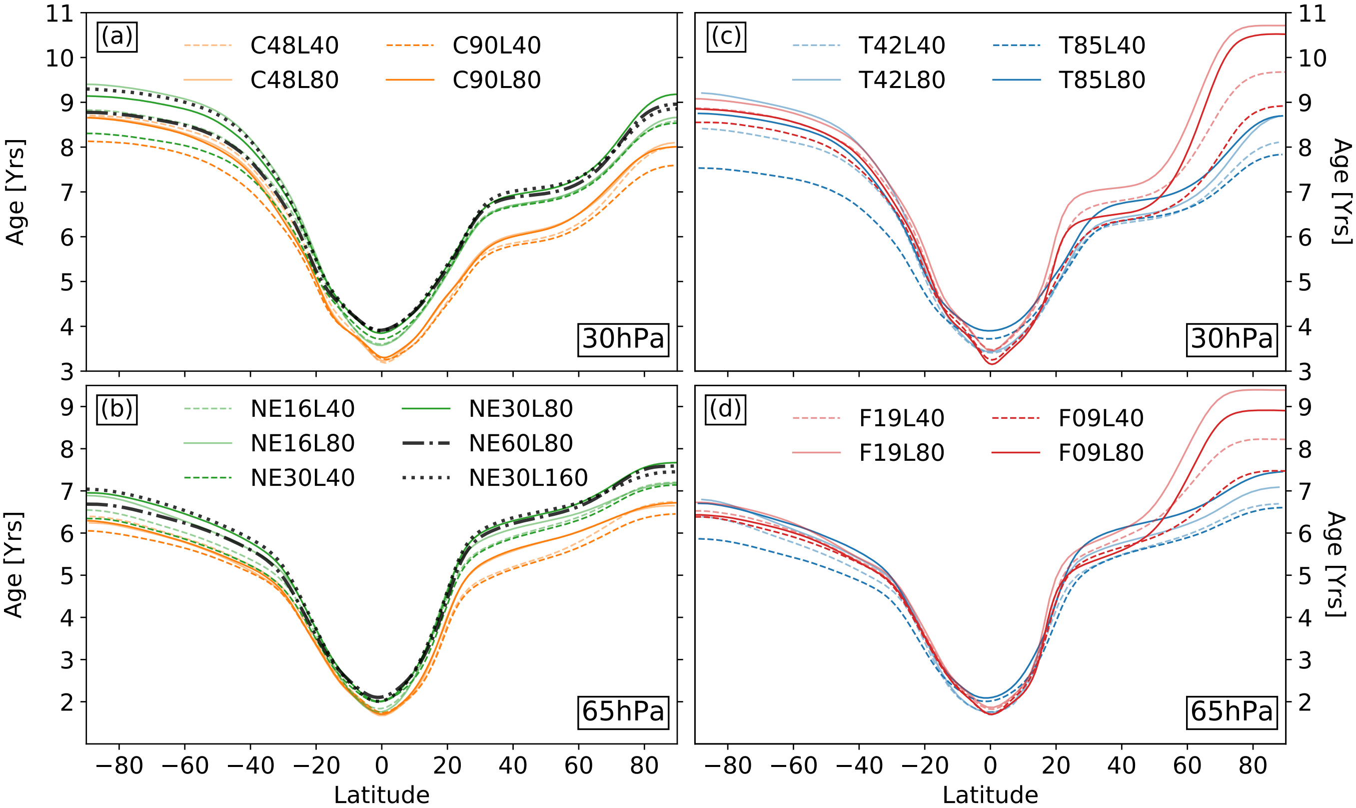

Figure 11. Age of air (in years) from the specified (SP) integrations from all four dycores at all four resolutions are shown. (a) and (b) show the age for GFDL-FV3 and CAM-SE at (a) 30hPa and (b) 65hPa. (c) and (d) show the age for GFDL-PS and CAM-FV at (c) 30hPa and (d) 65hPa. The low (high) horizontal resolution runs are shown using lighter (dark) curves and the low (high) vertical resolution runs are shown using dashed (solid) curves. The GFDL-pseudospectral is shown in blue, the GFDL-FV3 in orange, the CAM-SE in green and the CAM-FV in red. Higher horizontal (0.5° L80; NE60L80) and higher vertical (1° L160; NE30L160) CAM-SE runs are shown using dashed-dotted and dotted black curves in (a) and (b) respectively.

Forcing	**Test 1: Free running (FR)**	**Test 2: Specified tropical winds (SP)**
Diabatic forcing	Eqns (1), (A1) and (A2) in Polvani and Kushner (2002)	
Planetary Boundary Layer	Rayleigh friction as given in Held and Suarez (1994), Page 1826, 1st equation in the box	
Topography	3 km amplitude sinusoidal wave 2 mountain given by Eqn (1) in Gerber and Polvani (2009)	
Model Top	Stratospheric sponge as prescribed in Polvani and Kushner (2002), para [24]	
Tracer forcing	Linearly increasing clock tracer with source region $p \geq 700hPa$, Eqn 2	
Tropical stratosphere	Undamped	Zonal winds damped as specified in Eqn 5

Table 1. A summary of the model configuration used in our test in order to ensure a more accurate model stratosphere representation, induce lower boundary wave forcing and to initialize idealized clock tracer.

Numerics	Resolution	Model Resolution	Output Grid
GFDL - pseudospectral (GFDL-PS)	T42L40	$2.8^\circ \times 2.8^\circ$	$128 \times 64 \times 40$
	T42L80		$128 \times 64 \times 80$
	T85L40	$1.4^\circ \times 1.4^\circ$	$256 \times 128 \times 40$
	T85L80		$256 \times 128 \times 80$
GFDL - cubed sphere finite volume (GFDL-FV3)	C48L40	$2^\circ \times 2^\circ$	$192 \times 96 \times 40$
	C48L80		$192 \times 96 \times 80$
	C90L40	$1^\circ \times 1^\circ$	$360 \times 96 \times 40$
	C90L80		$360 \times 180 \times 80$
CAM - spectral element (CAM-SE)	NE16np4L40	$2^\circ \times 2^\circ$	$256 \times 129 \times 40$
	NE16np4L80		$256 \times 129 \times 40$
	NE30np4L40	$1^\circ \times 1^\circ$	$512 \times 257 \times 80$
	NE30np4L80		$512 \times 257 \times 80$
CAM - finite volume (CAM-FV)	F19L40	$2.5^\circ \times 1.9^\circ$	$144 \times 96 \times 40$
	F19L80		$144 \times 96 \times 40$
	F09L40	$1.25^\circ \times 0.9^\circ$	$288 \times 192 \times 80$
	F09L80		$288 \times 192 \times 80$

Table 2. A summary of the dynamical cores compared in this study, and the numerical resolutions considered. Details on the "nomenclature" (e.g., T42 vs. C48) are provided in Appendix B. Identical 40 (80) vertical levels, as specified in Polvani and Kushner (2002), were chosen for all cores. For the GFDL-PS and CAM-SE, the output grid only reflects the grid to which the output was interpolated, and does not provide an accurate representation of the model's effective resolution. For the GFDL-PS core, the model resolution is more accurately represented by the number of resolved spherical harmonics (up to wavenumber 42 or 85). For GFDL-FV3 and CAM-SE, the "true" resolution is determined by the number of points on each face of the cube. Thus, the NE16np4 (NE30np4) grid is identical in resolution to the C48 (C90) grid.

Numerics	**Horizontal Tracer Advection**	**Vertical Tracer Advection**
GFDL-**P**seudo**s**pectral	Default Spectral Scheme	Finite Volume Parabolic
GFDL-Cubed Sphere **F**inite **V**olume (GFDL-FV3)	Positive Definite Scheme as in Lin and Rood (1996) with Huynh constraint	Lin (2004) Vertically Lagrangian scheme. Remapping by monotonic & conservative cubic splines.
CAM-**S**pectral **E**lement	Kinmar & Gray RK4-SSP as in Kinnmark and Gray (1984)	Lin (2004) Vertically Lagrangian scheme. Remapping by Zerroukat monotonic splines.
CAM-**F**inite **V**olume	*Enhanced* Piecewise Parabolic Method (PPM) (Colella and Woodward 1984; Carpenter et al. 1990)	*Enhanced* Piecewise Parabolic Method (PPM) (Colella and Woodward 1984; Carpenter et al. 1990)

Table 3. Horizontal and vertical tracer advection schemes employed by the four dynamical cores considered in the study.